

QC

A85

V. 7

1954

SCIENCE

Digitized by the Internet Archive
in 2024

AUSTRALIAN JOURNAL OF PHYSICS

VOLUME 7

MELBOURNE

1954

AUSTRALIAN JOURNAL OF PHYSICS

Published by the Commonwealth Scientific and Industrial Research Organization, in collaboration with the Australian National Research Council and the Institute of Physics (Australian Branch), and under the general direction of the Editorial Board (Dr. N. S. Noble (Chairman), Professor J. S. Anderson, Professor Sir Macfarlane Burnet, Professor L. H. Martin, and Professor J. G. Wood). Volumes 1 to 5 of the Australian Journal of Physics and the Australian Journal of Chemistry issued as the Australian Journal of Scientific Research, Series A: Physical Sciences.

Issued quarterly, 30/- per annum

EDITORIAL ADVISORY COMMITTEE

Chairman and Editor: Dr. N. S. Noble

Members: Dr. G. H. Briggs

Professor L. H. Martin

Professor H. C. Webster

All enquiries and manuscripts should be forwarded to:

The Editor,

Australian Journal of Physics,

Commonwealth Scientific and Industrial Research Organization,

314 Albert Street, East Melbourne, C.2, Victoria.

CONTENTS

NUMBER 1, MARCH 1954

	PAGE
An Interpretation of the Field Tensor in the Unified Field Theory. By N. W. Taylor	1
Twisted Magnetic Fields in Conducting Fluids. By J. W. Dungey and R. E. Loughhead	5
On a Theory due to I. Fényes. By A. F. Nicholson	14
Some Electron Collision Cross Sections of CaII. By J. T. Jefferies ..	22
The Continuous Radiative Absorption Cross Section of FeXIV and the Coronal Temperature. By Alma Werner	25
Space Charge Wave Amplification in a Shock Front and the Fine Structure of Solar Radio Noise. By Hari K. Sen	30
On the Criterion for Dielectric Breakdown in Ionic Crystals. By J. J. O'Dwyer	36
The Integral and Differential Range Spectra of Sea-level Mesons. By A. J. Dyer	49
The Lattice Component of the Thermal Conductivity of Metals and Alloys. By P. G. Klemens	57
The Thermal Conductivity of Pure Metals at Low Temperatures According to the Free Electron Theory. By P. G. Klemens	64
The Electrical and Thermal Conductivities of Monovalent Metals. By P. G. Klemens	70
X-ray Line Broadening and Pure Diffraction Contours. By R. I. Garrod, J. F. Brett, and J. A. Macdonald	77
Galactic Radiation at Radio Frequencies. VII. Discrete Sources with Large Angular Widths. By J. G. Bolton, K. C. Westfold, G. J. Stanley, and O. B. Slee	96
Galactic Radiation at Radio Frequencies. VIII. Discrete Sources at 100 Mc/s between Declinations $+50^\circ$ and -50° . By J. G. Bolton, G. J. Stanley, and O. B. Slee	110
Observations of the General Background and Discrete Sources of 18.3 Mc/s Cosmic Noise. By C. A. Shain and C. S. Higgins	130
A Comparison of the Intensities of Cosmic Noise Observed at 18.3 Mc/s and at 100 Mc/s. By C. A. Shain	150
High Multiple Radio Reflections from the F_2 Layer of the Ionosphere at Brisbane. By K. Baird	165
Convection from a Large Horizontal Surface. By C. H. B. Priestley ..	176
Vertical Heat Transfer from Impressed Temperature Fluctuations. By C. H. B. Priestley	202
Deviations from Matthiessen's Rule for Cold-drawn Wires. By G. J. Ogilvie and W. K. Clothier	210

NUMBER 2, JUNE 1954

	PAGE
The Multiple Scattering of Electrons and Positrons. By C. B. O. Mohr and L. J. Tassie	217
Recent Studies of Chromospheric Spicules. By J. H. Rush and W. O. Roberts	230
On the Choice of Glasses for Cemented Achromatic Aplanatic Doublets. By W. H. Steel	244
On the Optical Properties of Components for Birefringent Filters. By R. G. Giovanelli and J. T. Jefferies	254
Dipole Resonant Modes of an Ionized Gas Column. By R. E. B. Makinson and D. M. Slade	268
The Decay of the 7.68 MeV State in ^{12}C . By R. G. Uebergang ..	279
The Response of a Sodium Iodide Scintillation Counter to 18 MeV γ -Radiation. By J. G. Campbell and A. J. F. Boyle	284
The Measurement and Reduction of Distortion in Thick Emulsions. By V. D. Hopper, Y. K. Lim, and Madeline C. Walters.. ..	288
Observations of the 21 cm Line from the Magellanic Clouds. By F. J. Kerr, J. F. Hindman, and B. J. Robinson	297
The Barometer Coefficient and Air Mass Effects on Cosmic Rays at Macquarie Island. By R. M. Jacklyn	315
An Investigation of Finite Strain in an Isotropic Material Subjected to Hydrostatic Pressure and its Seismological Applications. By A. Keane	322
On the Secular Variation of Rainfall at Adelaide. By E. A. Cornish ..	334
<i>Short Communications</i>	
The Capacitance of an Anchor Ring. By T. S. E. Thomas	347
On the Photodisintegrations ${}^6\text{Li}(\gamma, d){}^4\text{He}$ and ${}^6\text{Li}(\gamma, t){}^3\text{He}$. By E. W. Titterton and T. A. Brinkley	350
Delayed Disintegration of Heavy Fragments. By Francoise A. Brisbout and V. D. Hopper	352
Comments on a Paper by E. G. Bowen entitled "The Influence of Meteoritic Dust on Rainfall". 1. By W. C. Swinbank	354
Comments on a Paper by E. G. Bowen entitled "The Influence of Meteoritic Dust on Rainfall". 2. By D. F. Martyn	358
Corrigendum	364

NUMBER 3, SEPTEMBER 1954

	PAGE
On the Coulomb and Hulthén Potentials. By S. T. Ma	365
Electron Screening and Thermonuclear Reactions. By E. E. Salpeter..	373
Optical Diffraction Effects Produced by Amplitude and Phase Changes in the Wave Front. By G. H. Godfrey	389
The Influence of the Cathode Material on Measured Breakdown Strengths of Solid and Liquid Dielectrics. By J. J. O'Dwyer	400
The Residual Range of Delayed Particles in Extensive Air Showers. By V. C. Officer and P. J. Eccles	410
Momentum Distribution and Charge Ratio of μ -Mesons at Zenith Angles in the East-West Plane. By J. R. Moroney and J. K. Parry	423
Harmonics in the Spectra of Solar Radio Disturbances. By J. P. Wild, J. D. Murray, and W. C. Rowe.. .. .	439
Observations of Cosmic Noise at 9.15 Mc/s. By C. S. Higgins and C. A. Shain	460
The Association of Pulsating and Flaming Auroras with Complete Iono- spheric Absorption at Macquarie Island. By G. Major	471
Variations of Intensity of the Aurora at Macquarie Island. By F. Jacka	477
Programme Design for the C.S.I.R.O. Mark I Computer. III. Adaptation of Routines for Elaborate Arithmetical Operations. By T. Pearcey and G. W. Hill	485
The Effect of Interpretive Techniques on Functional Design of Computers. By T. Pearcey, G. W. Hill, and R. D. Ryan	505

Short Communications

The Contribution of Phonons to the Thomson Coefficient. By P. G. Klemens	520
Fluctuations of Long-Period Accumulations of Daily Rainfall Amounts. By J. Neumann	522
Polarization of Bremsstrahlung. By E. G. Muirhead and K. B. Mather..	527

NUMBER 4, DECEMBER 1954

	PAGE
The Multivariate t -Distribution Associated with a Set of Normal Sample Deviates. By E. A. Cornish	531
The Angular Distribution of High Energy Electrons in Air Showers. I. Landau Approximation. By M. H. Kalos and J. M. Blatt ..	543
The Albedo for the Atomic Scattering of Optical Radiation. By R. G. Giovanelli and J. T. Jefferies	570
The Emission of Radiation from Model Hydrogen Chromospheres. II. By J. T. Jefferies and R. G. Giovanelli	574
The Multiple Scattering of Protons in Nuclear Emulsions. By J. R. Bird and K. C. Hines	586
Measurements of the Cosmic Ray Neutron Rate in the Himalayas and Australian Alps. By K. B. Mather	601
Aerial Smoothing in Radio Astronomy. By R. N. Bracewell and J. A. Roberts	615
The Attenuation of Light by Meteoric Dust in the Upper Atmosphere. By R. G. Giovanelli	641
<i>Short Communications</i>	
Sea Surface Temperatures. By F. K. Ball.. .. .	649
Microinterferometric Examination of Nuclear Emulsion Plates. By B. S. Thornton	652
The Noise Generated in a Coil with a Ferromagnetic Core. By G. Builder and D. Haneman	654
Scattering of 4.9 MeV Protons by ^{27}Al . By K. B. Mather	658
Index to Volume 7	661

AN INTERPRETATION OF THE FIELD TENSOR IN THE UNIFIED FIELD THEORY

By N. W. TAYLOR*

[Manuscript received September 3, 1953]

Summary

It is assumed that the skew symmetric part of the field tensor g_{ik} is a complex, self-dual tensor. This permits the whole set of field equations for free space to be derived directly from the theory without the introduction of an electric current density tensor. However, with this assumption it appears impossible for spherically symmetric electric and magnetic fields to exist in free space.

I. THE FIELD EQUATIONS

It has been shown (Taylor 1952) how the complete set of Maxwell's electromagnetic equations in free space can be expressed simply by

$$\frac{\partial}{\partial x^k} (F^{ik}) = 0,$$

where F^{ik} is the tensor density derived from a complex, self-dual tensor F_{ik} . Hence, if the skew symmetric part of the field tensor g_{ik} is supposed to be of the same type as this F_{ik} , then the system of equations (I) of Einstein (1951, p. 138) gives the complete description of the field for the case of free space, without the insertion of an additional equation defining the current and charge density.

The derivation of Einstein's system (I) was a generalization of the process of obtaining the field equations for empty space in the purely gravitational field. Also, the system (I) reduces to these equations when the skew symmetric part of g_{ik} vanishes. Hence it appears as an advantage of the present interpretation of the skew symmetric part of g_{ik} that the current and charge vector density is zero.

II. THE SELF-DUAL FIELD TENSOR

The condition for self-duality of the skew symmetric part of g_{ik} requires special investigation in the present theory. Write

$$g_{ik} = a_{ik} + \varphi_{ik}, \quad \dots \quad (1)$$

where a_{ik} is the symmetric part and φ_{ik} the skew symmetric part of g_{ik} . In the electromagnetic field theory the restriction on F_{ik} becomes

$$\sqrt{-a} \alpha^{ik} \alpha^{ls} F_{rs} = \frac{1}{2} \epsilon^{iklm} F_{lm}, \quad \dots \quad (2)$$

In the present theory, however, tensor densities may be formed using the determinant g as well as a , and suffixes may be raised by a^{ik} , g^{ik} , or g^{ik} (the symmetric

* Department of Mathematics, University of New England, Armidale, N.S.W.

part of g^{ik}). The various possible combinations give many equations restricting the φ_{ik} , all of which could be accepted as generalizations of (2). However, it can be shown by direct calculation that in the case of the spherically symmetric field described in Section III, each of these combinations gives an inconsistent set of equations, except the three

$$\sqrt{a}a^{ir}a^{ks}\varphi_{rs}=\frac{1}{2}\varepsilon^{iklm}\varphi_{lm}, \quad \dots\dots\dots (3)$$

$$\sqrt{g}g^{ir}g^{ks}\varphi_{rs}=\frac{1}{2}\varepsilon^{iklm}\varphi_{lm}, \quad \dots\dots\dots (4)$$

$$\sqrt{g}g^{ri}g^{sk}\varphi_{rs}=\frac{1}{2}\varepsilon^{iklm}\varphi_{lm}. \quad \dots\dots\dots (5)$$

The first of these involves only three distinct equations and obviously does not impose too rigid a restriction even should an arbitrary field be under discussion. The equations (4) and (5) are not linear in the φ_{ik} as they stand, and their investigation in the arbitrary field would be much more complicated.

It is possible to show that (3) is a solution to (4) and (5), however. This can be done using formulae given by Kurşunoğlu (1952) adjusted to suit the present notation. These are

$$g=a(1+\Omega+\Lambda^2), \quad \dots\dots\dots (6)$$

where

$$\Omega=\frac{1}{2}\varphi_{ik}\varphi^{ik}, \quad \dots\dots\dots (7)$$

and

$$\Lambda=\frac{1}{4}f^{ik}\varphi_{ik}, \quad \dots\dots\dots (8)$$

with

$$f^{ik}=\varepsilon^{iklm}\varphi_{lm}/2\sqrt{a}. \quad \dots\dots\dots (9)$$

Also

$$g_{ik}^{\pm}=a^{ik}-g_{\pm}^{ij}\varphi_{ij}^k, \quad \dots\dots\dots (10)$$

and

$$g_{ik}^{\pm}=(\varphi^{ik}+\Lambda f^{ik})/(1+\Omega+\Lambda^2). \quad \dots\dots\dots (11)$$

There is also the identity

$$\delta_k^i\Lambda=f^{il}\varphi_{kl}. \quad \dots\dots\dots (12)$$

All indices (except those of g^{ik} which has its usual definition) are raised by a^{ik} in the formulae (6)–(12).

Since it is assumed that (3) holds, (9) reduces to

$$f^{ik}=\varphi^{ik}. \quad \dots\dots\dots (13)$$

From the equations (7) and (8) we then have $\Lambda=\frac{1}{2}\Omega$, and so (11) simplifies to

$$g_{ik}^{\pm}=\varphi^{ik}/(1+\frac{1}{2}\Omega). \quad \dots\dots\dots (14)$$

(10) and (11) then give

$$g_{ik}^{\pm}=g_{\pm}^{ik}+g_{\pm}^{ik}=a^{ik}+(\varphi^{ik}-\varphi^{il}\varphi_{il}^k)/(1+\frac{1}{2}\Omega).$$

Since (12) reduces to

$$\frac{1}{2}\delta_k^i\Omega=\varphi^{il}\varphi_{kl}, \quad \dots\dots\dots (15)$$

this becomes

$$g_{ik}=(a^{ik}+\varphi^{ik})/(1+\frac{1}{2}\Omega). \quad \dots\dots\dots (16)$$

Consider (4). Since (6) is now $g=a(1+\frac{1}{2}\Omega)^2$, the left-hand side is

$$\sqrt{a}(a^{ir}+\varphi^{ir})(a^{ks}+\varphi^{ks})\varphi_{rs}/(1+\frac{1}{2}\Omega),$$

that is,

$$\sqrt{a}(\varphi^{ik}+\varphi^{ir}\varphi_r{}^k+\varphi^{ks}\varphi_s{}^i+\varphi^{ir}\varphi^{ks}\varphi_{rs})/(1+\frac{1}{2}\Omega),$$

that is,

$$\sqrt{a}(\varphi^{ik}+\varphi^{ir}\varphi^{ks}\varphi_{rs})/(1+\frac{1}{2}\Omega).$$

Using (15) this expression simplifies to $\sqrt{a}\varphi^{ik}$. By (3), this is equal to $\frac{1}{2}\varepsilon^{iklm}\varphi_{lm}$, which is also the right-hand side of (4). Hence (4) is satisfied by (3).

It can be shown at once also that if (3) holds then (5) is satisfied. For, the left-hand side of (5) is

$$\sqrt{a}(a^{ir}-\varphi^{ir})(a^{ks}-\varphi^{ks})\varphi_{rs}/(1+\frac{1}{2}\Omega),$$

that is,

$$\sqrt{a}(\varphi^{ik}+\varphi^{ir}\varphi^{ks}\varphi_{rs})/(1+\frac{1}{2}\Omega),$$

as before.

III. THE SPHERICALLY SYMMETRIC FIELD

Papapetrou (1948) has found that the general static spherically symmetric g_{ik} can be written in the form

$$g_{ik} = \left. \begin{array}{cccc} -\alpha & 0 & 0 & w \\ 0 & -\beta & r^2 v \sin \theta & 0 \\ 0 & -r^2 v \sin \theta & -\beta \sin^2 \theta & 0 \\ -w & 0 & 0 & \gamma \end{array} \right\} \dots \dots (17)$$

It is necessary to determine now the restriction imposed on these components by the condition that the skew symmetrical part is self dual. In this case it is easily seen by straightforward calculation that (4) and (5) have just one solution, that given by (3), namely,

$$w = i \sqrt{\frac{\alpha \gamma}{\beta^2}} r^2 v = i \sqrt{\frac{\alpha \gamma}{\beta^2}} f, \dots \dots \dots (18)$$

using $f=r^2v$ (Bonnor 1951, 1952).

Without this restriction the form (17) would be sufficiently general to permit the description of the spherically symmetric field in a dielectric and magnetic medium (Taylor 1953). However, for the reasons given in Section I, the present theory does not include such media.

The terms w and f are complex. The electric field components are now not distinguished from the magnetic by position in the array (17). The separation is made by resolving the w and f into real and imaginary parts, there being freedom of choice, until an electric current vector is defined, as to which part corresponds to the electric and which to the magnetic field.

Since w and f cannot now be made to vanish separately, it is necessary to use the completely general static spherically symmetric solution of Einstein's (1951) system of equations (I) given by Bonnor (1952). In the following analysis the references will be to this paper.

Substituting for γ from (18) into Bonnor's equation (2.21),

$$(f^2 + \beta^2)w^2/\alpha = -cf^2. \quad \dots\dots\dots (19)$$

According to Bonnor's equation (2.20) the expression on the left-hand side of (19) is also equal to l^2 . Hence

$$f = \pm il/\sqrt{c} = \text{a constant}. \quad \dots\dots\dots (20)$$

From (20) and (2.2) of Bonnor

$$B = f\beta'/(\beta^2 + f^2),$$

and so Bonnor's equations (2.24) and (2.25) become, using this and (20),

$$\beta'' - (\alpha'/2\alpha)\beta' - 2\alpha(\beta^2 - f^2)/(\beta^2 + f^2) = 0, \quad \dots\dots\dots (21)$$

and

$$(\beta')^2 - 4\alpha\beta = 0. \quad \dots\dots\dots (22)$$

Eliminating α from (21) and (22),

$$\beta'' - (\beta''/\beta' - \beta'/2\beta)\beta' - \beta'^2(\beta^2 - f^2)/[2\beta(\beta^2 + f^2)] = 0.$$

This simplifies to

$$\beta' = 0,$$

which, together with (20), leads to inconsistencies in the field equations. For example, R_{22} , given by Bonnor's equations (2.8) and (2.2), cannot be made to vanish.

The present interpretation of g_{ik} would imply the impossibility of the existence of static spherically symmetric electromagnetic fields when the field equations are Einstein's system (I).

IV. ACKNOWLEDGMENT

The author is indebted to the referee for valuable discussion and assistance in the development of this paper.

V. REFERENCES

- BONNOR, W. B. (1951).—Static spherically symmetric solutions in Einstein's unified field theory. *Proc. Roy. Soc. A* **209**: 353-68.
- BONNOR, W. B. (1952).—The general static spherically symmetric solution in Einstein's unified field theory. *Proc. Roy. Soc. A* **210**: 427-34.
- EINSTEIN, A. (1951).—"The Meaning of Relativity." 5th Ed. Appendix II. (Methuen: London.)
- KURŞUNOĞLU, B. (1952).—Gravitation and electrodynamics. *Phys. Rev.* **88**: 1369.
- PAPAPETROU, A. (1948).—Static spherically symmetric solutions in the unitary field theory. *Proc. R. Irish Acad. A* **52**: 69-86.
- TAYLOR, N. W. (1952).—A simplified form of the relativistic electromagnetic equations. *Aust. J. Sci. Res. A* **5**: 423-9.
- TAYLOR, N. W. (1953).—The relativistic electromagnetic equations in a material medium. *Aust. J. Phys.* **6**: 1-9.

TWISTED MAGNETIC FIELDS IN CONDUCTING FLUIDS

By J. W. DUNGEY* and R. E. LOUGHHEAD†

[Manuscript received September 3, 1953]

Summary

The formation of loops in the lines of force of a twisted magnetic field confined within a cylinder of radius R , first suggested by Alfvén (1950*a*), is discussed by the method of normal modes. The model first becomes unstable with respect to modes which do not lead to the formation of loops. Ignoring this, the condition obtained for loop formation is that the pitch of the twisted field be less than πR .

The velocity of Alfvén waves in this model is also discussed.

I. INTRODUCTION

There are several reasons why the study of twisted magnetic fields should be relevant to solar physics. For example, some prominences show filaments which give the impression of being twisted. Alfvén (1950*a*) drew attention to the formation of loops in a string when it is twisted beyond a certain limit, and pointed out that this will also happen to a magnetic field. While Alfvén was concerned with the origin of the terrestrial and solar magnetic fields, this phenomenon might also be important in prominences and perhaps in the generation of sunspots, although it is not involved in Alfvén's (1950*b*) own theory of sunspots.

Alfvén (1950*a*) discussed the model in which a tube of force of radius R containing a uniform magnetic field is given a uniform twist. By consideration of the magnetic energy he showed that the condition for loop formation is approximately

$$R > (\sqrt{5} - 1)p, \dots\dots\dots (1)$$

where $2\pi p$ is the pitch of the twist. Lundquist (1951) investigated the stability of the general twisted field, which is represented in cylindrical coordinates by the two components H_ϕ and H_z , these being arbitrary functions of r . He found the field to be unstable if

$$\int H_\phi^2 r dr > 2 \int H_z^2 r dr,$$

which for Alfvén's model becomes

$$R > 2p. \dots\dots\dots (2)$$

However, this result is obtained by considering the stability with respect to a specific type of displacement which, as Lundquist points out, is not the most critical. In the present paper the stability of Alfvén's model is discussed by the method of normal modes.

* School of Physics, University of Sydney.

† School of Physics, University of Sydney ; present address : Division of Physics, C.S.I.R.O., University Grounds, Sydney.

II. THE NORMAL MODES

We first formulate the basic magneto-hydrodynamic equations in the case of a general twisted field

$$\mathbf{H} = (0, H_\varphi(r), H_z(r)), \quad \dots\dots\dots (3)$$

contained wholly within a cylindrical tube of force of radius R . From a consideration of small oscillations of frequency ω in the conducting medium, we obtain the criterion for stability by putting $\omega=0$. It can be shown that the stability does not depend on the density or compressibility of the conducting fluid, which is thus taken to have uniform mass density μ and to be incompressible. As usual, the electrical conductivity is taken to be infinite. A small material velocity \mathbf{v} then causes a small change \mathbf{h} in the magnetic field \mathbf{H} determined by

$$\frac{\partial \mathbf{h}}{\partial t} = (\mathbf{H} \cdot \text{grad}) \mathbf{v} - (\mathbf{v} \cdot \text{grad}) \mathbf{H}. \quad \dots\dots\dots (4)$$

The effect of the magnetic force on the motion is represented to the first order by

$$4\pi\mu \frac{\partial \mathbf{v}}{\partial t} = -\text{grad } \psi + (\mathbf{H} \cdot \text{grad}) \mathbf{h} + (\mathbf{h} \cdot \text{grad}) \mathbf{H}, \quad \dots\dots\dots (5)$$

where $\psi/4\pi$ is the variation in "total pressure", that is, gas pressure plus magnetic pressure ($=H^2/8\pi$).

When $\omega \neq 0$ equations (4) and (5) may be solved with the aid of the relation

$$\text{div } \mathbf{v} = 0. \quad \dots\dots\dots (6)$$

By seeking solutions proportional to

$$\exp i(\omega t + kz + m\varphi), \quad \dots\dots\dots (7)$$

and writing

$$K = kH_z + \frac{m}{r}H_\varphi, \quad \dots\dots\dots (8)$$

we obtain from (4), (5), and (6) the following equations:

$$i\omega h_r = iKv_r, \quad \dots\dots\dots (9)$$

$$i\omega h_\varphi = iKv_\varphi - v_r \frac{\partial H_\varphi}{\partial r} + v_r \frac{H_\varphi}{r}, \quad \dots\dots\dots (10)$$

$$i\omega h_z = iKv_z - v_r \frac{\partial H_z}{\partial r}, \quad \dots\dots\dots (11)$$

$$4\pi\mu i\omega v_r = -\frac{\partial \psi}{\partial r} + iKh_r - 2\frac{H_\varphi}{r}h_\varphi, \quad \dots\dots\dots (12)$$

$$4\pi\mu i\omega v_\varphi = -im\psi/r + iKh_\varphi + \left(\frac{\partial H_\varphi}{\partial r} + \frac{H_\varphi}{r}\right)h_r, \quad \dots\dots\dots (13)$$

$$4\pi\mu i\omega v_z = -ik\psi + iKh_z + \frac{\partial H_z}{\partial r}h_r, \quad \dots\dots\dots (14)$$

$$\frac{\partial v_r}{\partial r} + \frac{v_r}{r} + imv_\varphi/r + ikv_z = 0. \quad \dots\dots\dots (15)$$

After some elimination, we then find

$$(4\pi\mu\omega^2 - K^2) \left(\frac{\partial v_r}{\partial r} + \frac{v_r}{r} \right) + \frac{2mKH_\varphi}{r^2} v_r = i\omega \left(k^2 + \frac{m^2}{r^2} \right) \psi, \quad \dots \quad (16)$$

$$\omega \frac{\partial \psi}{\partial r} + i(4\pi\mu\omega^2 - K^2) v_r + 2 \frac{H_\varphi}{r} \left\{ i v_r \frac{\partial H_\varphi}{\partial r} - \frac{i H_\varphi v_r}{r} \left(\frac{4\pi\mu\omega^2 + K^2}{4\pi\mu\omega^2 - K^2} \right) - \frac{\omega K m \psi}{(4\pi\mu\omega^2 - K^2)r} \right\} = 0. \quad \dots \quad (17)$$

From these equations the necessary boundary conditions for the components v_r and ψ at $r=R$ readily follow; for it is evident from the form of equation (16), and is in any case physically obvious, that v_r must be continuous across the boundary $r=R$. Also, the discontinuous change $\delta\psi$ in ψ across the boundary is related to the value of v_r at $r=R$ through the term involving $\partial H_\varphi / \partial r$ in (17), which yields

$$v_r = i\omega \cdot \frac{R}{H_\varphi^2} \delta\psi. \quad \dots \quad (18)$$

In Alfvén's model, where H_z is constant and H_φ takes the form

$$H_\varphi = Ar, \quad \dots \quad (19)$$

appropriate to a uniformly twisted field,

$$K = kH_z + mA$$

is a constant for $r < R$ and vanishes for $r > R$. Eliminating v_r between equations (16) and (17) we obtain the modified Bessel equation

$$\frac{\partial^2 \psi}{\partial r^2} + \frac{1}{r} \frac{\partial \psi}{\partial r} - \left(k'^2 + \frac{m^2}{r^2} \right) \psi = 0, \quad \dots \quad (20)$$

where

$$k'^2 = k^2 \left\{ 1 - \frac{4K^2 A^2}{(4\pi\mu\omega^2 - K^2)^2} \right\}. \quad \dots \quad (21)$$

Then, since ψ remains finite at the origin, the variation of the total pressure inside the tube of force is given by

$$\psi = \alpha I_m(k'r) e^{i(\omega t + kz + m\varphi)}, \quad \dots \quad (22)$$

where $I_m(k'r)$ is the modified Bessel function (Jeffreys 1950) of the first kind of order m and α is an arbitrary constant.

Outside the tube of force, $k' = k$, and then the solution of (20), vanishing at infinity, is

$$\psi = \beta K_m(kr) e^{i(\omega t + kz + m\varphi)}, \quad \dots \quad (23)$$

where $K_m(kr)$ is the modified Bessel function of the second kind of order m and β is an arbitrary constant. Generalizing to allow for different densities μ_1 for

$r < R$ and μ_2 for $r > R$ and substituting from (22) and (23) we find that the boundary requirements at $r = R$ are satisfied if

$$\frac{(4\pi\mu_1\omega^2 - K^2) \frac{x}{I_m(x)} \frac{dI_m(x)}{dx} - 2mKA}{(4\pi\mu_1\omega^2 - K^2)^2 - 4K^2A^2} = \frac{y \cdot \frac{dK_m(y)}{dy}}{4\pi\mu_2\omega^2 K_m(y) - A^2 y \frac{dK_m(y)}{dy}}, \quad (24)$$

where $x = k'R$ and $y = kR$.

Before proceeding to the case $\omega = 0$ it is of interest to apply the above general formulae to the propagation of Alfvén waves along a uniform untwisted tube of force ($A = 0$).

III. ALFVÉN WAVES IN A UNIFORM UNTWISTED TUBE OF FORCE

In this case the velocity of torsional waves in an infinite conducting fluid of density μ_1 follows from the fundamental equations (9)–(15) by putting

$$v_r = 0. \quad (25)$$

Then $\psi = 0$ and $4\pi\mu_1\omega^2 - K^2 = 0$. Hence for the velocity of torsional waves we obtain

$$V = \pm V_a,$$

where V_a is Alfvén's velocity

$$V_a = \frac{H_z}{\sqrt{4\pi\mu_1}}. \quad (26)$$

If, however, we drop the condition (25) and seek instead solutions which satisfy the general boundary condition (24) we find that the solution for $r < R$ is now

$$\psi = \alpha I_m(kr) e^{i(\omega t + kz + m\phi)}.$$

Substitution in (24) yields

$$V_m = \pm V_a \left\{ 1 - \frac{\mu_2}{\mu_1} \cdot \frac{I_m^{-1}(y) \frac{dI_m(y)}{dy}}{K_m^{-1}(y) \frac{dK_m(y)}{dy}} \right\}^{-\frac{1}{2}}, \quad (27)$$

where $y = kR$ and V_m is the velocity of waves parallel to the z -axis for any given mode m . Equation (27) shows the reduction of wave velocity due to the inertia of the surrounding fluid. Simple consideration of the expansions of the Bessel functions for large and small y shows that

$$V_m \rightarrow \pm \frac{1}{\sqrt{2}} V_a$$

as kR becomes very large or very small. This conclusion is true for all m except $m = 0$, which is exceptional in that $V_m \rightarrow \pm V_a$ for small kR . The variation of wave velocity with kR is illustrated in Figure 1 for the case $m = 1$, with $\mu_1 = \mu_2$.

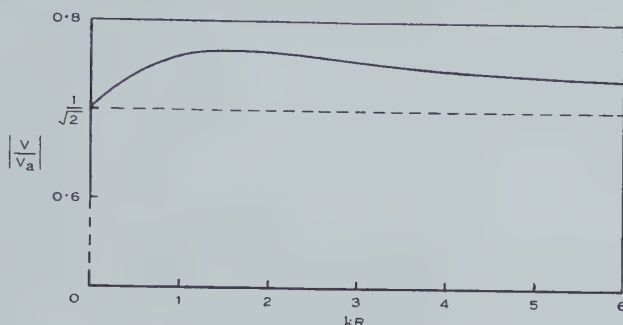


Fig. 1

IV. THE CRITICAL CASE $\omega=0$

When $\omega=0$, equation (4) requires $v=0$, so that the fundamental equation (5) must now be supplemented by the relation

$$\operatorname{div} \mathbf{h}=0, \quad \dots\dots\dots (28)$$

which, when $\omega \neq 0$, is a consequence of (4). We thus obtain the equations

$$\begin{aligned} \frac{\partial \psi}{\partial r} &= iK h_r - \frac{2H_\phi}{r} \cdot h_\phi, \\ \frac{im\psi}{r} &= iK h_\phi + \left(\frac{\partial H_\phi}{\partial r} + \frac{H_\phi}{r} \right) h_r, \\ ik\psi &= iK h_z + \frac{\partial H_z}{\partial r} \cdot h_r, \\ \frac{\partial h_r}{\partial r} + \frac{h_r}{r} + imh_\phi/r + ikh_z &= 0, \end{aligned}$$

which, if $H_\phi = Ar$, reduce to the forms

$$K \left(\frac{\partial h_r}{\partial r} + \frac{h_r}{r} \right) - \frac{2mA}{r} h_r + i \left(k^2 + \frac{m^2}{r^2} \right) \psi = \frac{\partial K}{\partial r} h_r \quad \dots\dots\dots (29)$$

and

$$K \frac{\partial \psi}{\partial r} + \frac{2mA}{r} \psi - i(K^2 - 4A^2) h_r = -ir \frac{\partial A^2}{\partial r} h_r. \quad \dots (30)$$

With Alfvén's model, the right-hand sides of (29) and (30) vanish for $r < R$, and we can obtain by elimination Bessel's equation giving the solution

$$\psi = \alpha I_m(k'r) e^{i(\omega t + kz + m\phi)}, \quad \dots\dots\dots (31)$$

where

$$k'^2 = k^2 \left(1 - \frac{4A^2}{K^2} \right). \quad \dots\dots\dots (32)$$

The discontinuous change in ψ at $r=R$ can be deduced from the derivatives on the right-hand sides of (29) and (30) by noting that, for $r > R$, $K=A=0$

and hence $\psi=0$. Equation (29) implies the continuity of h_r/K at $r=R$, and then equation (30) shows that

$$\psi = -i \cdot R \cdot A^2 \cdot \frac{h_r}{K}. \quad \dots\dots\dots (33)$$

This is the boundary condition at $r=R$ and, with the aid of (30), may be written in the form

$$\left\{ \frac{r}{\psi} \cdot \frac{\partial \psi}{\partial r} \right\}_{r=R} = 4 - \left(\frac{K}{A} \right)^2 - \frac{2mA}{K}, \quad \dots\dots\dots (34)$$

provided that $K^2 \neq 4A^2$.

The latter condition may be further restricted; for if $K^2 > 4A^2$, k' is real and then

$$\frac{r}{I_m(k'r)} \cdot \frac{dI_m(k'r)}{dr} \geq |m|. \quad \dots\dots\dots (35)$$

However, with $K^2 > 4A^2$, the right-hand side of (34) is less than $|m|$ and hence there are no solutions with $K^2 > 4A^2$.

The possibility of a solution with $K^2 = 4A^2$ requires separate investigation. From (30) it follows that mA/K must be negative and ψ proportional to $r^{|m|}$. Then equation (29) yields

$$2Kh_r = -i\psi \left\{ \frac{|m|}{r} + \frac{k^2 r}{|m|+1} \right\},$$

which requires

$$iKh_r/\psi > 0$$

at $r=R$ and thus contradicts the boundary condition (34). Hence no such solution exists.

When $K^2 < 4A^2$ it is convenient to define

$$\kappa^2 = k^2 \left(\frac{4A^2}{K^2} - 1 \right), \quad \dots\dots\dots (36)$$

and then the solution for $r < R$ takes the more familiar form

$$\psi = \alpha J_m(\kappa r) e^{i(\omega t + kz + m\phi)}, \quad \dots\dots\dots (37)$$

which is now an oscillatory function of r . Since κ can be made large by making K small solutions of the form (37) must exist. It is sufficiently general to consider only positive values of k , H_z , and A provided that both positive and negative values of m are allowed. If

$$p = H_z/A, \quad \dots\dots\dots (38)$$

then $2\pi p$ is the pitch of the uniformly twisted field. With this notation the inequality $K^2 < 4A^2$ may be written

$$(kp + m)^2 < 4,$$

and this eliminates the modes with $m \geq 2$.

V. THE CONDITIONS OF STABILITY

The critical values of A for stability are obtained from the solutions with $\omega=0$.

Let

$$F_m(x) = \frac{x}{J_m(x)} \cdot \frac{dJ_m(x)}{dx}, \quad \dots\dots\dots (39)$$

and

$$g_m(kp) = 4 - (kp+m)^2 - \frac{2m}{kp+m}, \quad \dots\dots\dots (40)$$

then the boundary condition (34) is expressed by the relation

$$F_m(\kappa R) = g_m(kp). \quad \dots\dots\dots (41)$$

The following properties of the function $F_m(x)$ are subsequently required

$$\left. \begin{aligned} F_m(x) &= |m|, & \text{when } x=0, \\ \frac{dF_m(x)}{dx} &\leq 0, & \text{when } x \geq 0, \end{aligned} \right\} \dots\dots\dots (42)$$

and

$F_m(x)$ has poles at the zeros of $J_m(x)$.

For given values of m , k , and p equation (41) determines a set of values of R , the smallest of which will be called R_0 .

To determine the ranges of values of A for which imaginary eigenvalues of ω can occur we refer back to equation (20), which has solutions of the form

$$\psi = \alpha J_m(\kappa' r) e^{i(\omega t + k z + m \varphi)}, \quad \dots\dots\dots (43)$$

where

$$\kappa'^2 = k^2 \left\{ \frac{4K^2 A^2}{(4\pi\mu\omega^2 - K^2)^2} - 1 \right\}. \quad \dots\dots\dots (44)$$

When $K^2 > 4A^2$ a small imaginary value of ω will still leave κ' real and the value of $F_m(\kappa'R)$ given by (24) will vary only slightly. An eigenvalue of ω is therefore determined by $\kappa'R \approx \kappa R_0$ and since, when ω is imaginary, $|\kappa'| < |\kappa|$, the condition for instability is expressed by the inequality

$$R > R_0. \quad \dots\dots\dots (45)$$

It is evident that instability will occur for $m < 0$ when $|kp+m|$ is small enough since then κ is large and R_0 small. This type of instability is not relevant to loop formation. If $kp+m=0$, the resultant motion has the same symmetry as the twisted field. Only the mode $m=-1$ can be observed with a string; the string as a whole takes a helical form. It will be seen that this occurs before loop formation, so that a complete study would require an investigation of the new equilibrium configuration of the field, but this seems to be rather intractable. It may also be noted that this type of instability will not always occur for more general models; it is facilitated when the pitch of the twist is independent of r .

The only modes for which the axial line of force moves are those with $m = \pm 1$; this is easily seen by substituting the solution (43) into equation (17), which then shows that $v_r = 0$ at $r = 0$ unless $m = \pm 1$. The other modes are

therefore irrelevant to a discussion of loop formation. If the field is unstable with respect to $m = -1$ only, the tube of force as a whole takes up a helical configuration, but does not form loops. The type of motion required in forming loops occurs only when the field is simultaneously unstable with respect to both the modes $m = \pm 1$. This is therefore the best criterion for loop formation obtainable from the present restricted analysis.

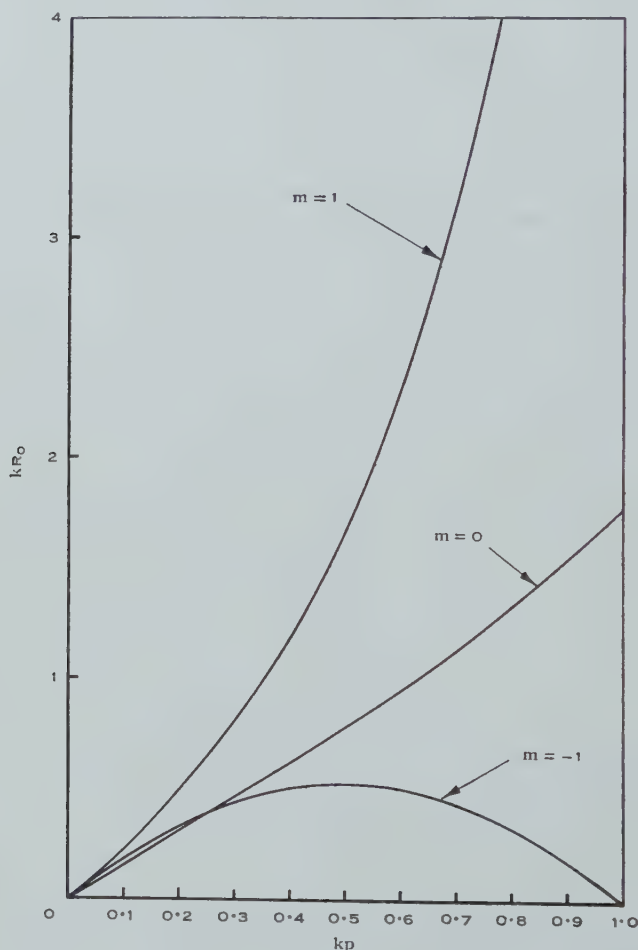


Fig. 2

The values of kR_0 obtained from (41) are plotted against kp in Figure 2 for $m = \pm 1$ and 0. It is clear from this diagram that, for any given kp , the corresponding value of kR_0 when $m=1$ exceeds the value when $m=-1$. Hence it follows from (45) that when the field is unstable with respect to $m=1$ it is always unstable with respect to $m=-1$, so that the condition for loop formation can be taken as the condition for instability with respect to $m=1$. For small values of kp this condition is

$$R > 2p, \dots\dots\dots (46)$$

which follows from the relation $R_0 \approx 2p$ implied by equation (41) when $m = \pm 1$ and kp is small. The criterion (46) is the same as that obtained by Lundquist (1951) (cf. equation (2)), and this is not altogether surprising, since the system is unstable with respect to many modes and Lundquist's disturbance may be a combination of these.

The graph also shows that loop formation cannot occur with $kp > 1$; this simply means that no more loops can be formed than there are twists. If there is more physical limitation on k , such that k must be an integral multiple of $2\pi/L$ say, the critical value of p for loop formation will be decreased; if, however, $\pi R \ll L$, the decrease will be only slight.

VI. CONCLUSIONS

We may summarize the above results as follows:

In the normal modes the disturbance varies as $\exp i(\omega t + kz + m\phi)$.

The velocity of Alfvén waves in a uniform untwisted tube of force is reduced by the motion in the surrounding material by a factor which approximates to $(1 + \mu_2/\mu_1)^{-\frac{1}{2}}$ for both long and short wavelengths (see Fig. 1).

A twisted tube of pitch $2\pi p$ is unstable with respect to a given mode when $|kp + m|$ is small enough; it is always stable when $|kp + m| > 2$. The formation of loops requires simultaneous instability with respect to two modes with $m = +1$ and -1 and with the same value of k . This requires $p < \frac{1}{2}R$.

VII. ACKNOWLEDGMENTS

The authors are indebted to the Research Committee, University of Sydney, for the grant of a Research Fellowship and a Research Studentship respectively.

VIII. REFERENCES

- ALFVÉN, H. (1950a).—*Tellus* **2**: 74.
 ALFVÉN, H. (1950b).—"Cosmical Electrodynamics." Ch. 5. (Oxford Univ. Press.)
 JEFFREYS, H., and JEFFREYS, B. S. (1950).—"Methods of Mathematical Physics." 2nd Ed. Ch. 21. (Cambridge Univ. Press.)
 LUNDQUIST, S. (1951).—*Phys. Rev.* **83**: 307.

ON A THEORY DUE TO I. FÉNYES

By A. F. NICHOLSON*

[*Manuscript received November 17, 1953*]

Summary

Fényes (1952) has attempted to base quantum mechanics on a theory of Markoff processes in the configuration space of any system considered.

It is shown here that for certain systems the allowable stationary-state solutions of the basic equations of this theory are more numerous than the stationary states predicted by quantum mechanics through the Schrödinger equation, and that it is not possible to identify the probability amplitude defined in this theory with the Schrödinger wave function. It is pointed out that the "total stochastic velocity" of a particle in this theory cannot equal the linear momentum divided by the mass, that Fényes' uncertainty relation is not equivalent to the Heisenberg relation, and that the total-stochastic-velocity operator has properties different from those of the quantum-mechanical momentum operator. It is shown that Fényes' proof that the addition of further parameters to a set of statistical parameters cannot lead to a set affording a causal description of the motion of a system is unsatisfactory, but that this conclusion may reasonably be drawn from his theory, by making a consistent auxiliary assumption.

I. INTRODUCTION

Fényes (1952) has attempted to describe the motion of a general microscopic system as a Markoff process in a configuration space, which is implied to be that of the Cartesian position coordinates of the particles constituting the system. Then quantum mechanics becomes a statistical theory. Only the briefest outline can be given here of Fényes' lengthy exposition.

The theory is based upon an analytical treatment of stochastic processes due to Kolmogoroff (1931, 1933). Fényes begins with a set of parameters (position coordinates) and probability and transition probability densities w and v which are functions of these parameters and the time. He takes these densities to be appropriately normalized, and interrelates them through an integral equation which defines a Markoff process. An integral relation then follows between the transition probability densities at three distinct instants of time. Fényes assumes the existence of certain limits in order to derive two differential equations from the original integral relations. One equation leads to a continuity or conservation equation, the Fokker equation. Fényes defines a "total stochastic velocity" and corresponding operators, and deduces an "uncertainty relation". He then further restricts the class of systems considered and defines a probability amplitude ψ . When expressed in terms of ψ , the Fokker equation is formally identical with the equation of "conservation of

* Physics Department, Birkbeck College, University of London; present address: Long Range Weapons Establishment, Salisbury, S.A.

probability" in quantum mechanics (see e.g. Tolman (1938)). Fényes modifies his original total stochastic velocity definition to define "the velocity corresponding to ψ ", which he claims is the same as the quantum-mechanical velocity.

After a consideration of formal analogies and further discussion, Fényes decides that the mathematical apparatus of quantum mechanics is identical with that of his treatment of Markoff processes, and that quantum-mechanical processes are special Markoff processes. This conclusion leads Fényes to suppose that further degrees of freedom, hidden parameters, must be assigned to the electron. Fényes dismisses von Neumann's (1932) objections to the existence of hidden parameters in quantum mechanics by stating that von Neumann proposes to reach a causal description in terms of ψ plus further parameters, and arguing that, since ψ characterizes the state of the system from the statistical standpoint, it cannot be considered as a causal state parameter.

Fényes then attempts to demonstrate that his set of statistical configuration coordinates cannot be completed by the addition of further parameters to a set of coordinates in terms of which a causal description of the motion may be given. He also tries to show that the state of a system for which the "scattertensor" does not vanish cannot be characterized by a phase-space distribution function of the kind considered in statistical mechanics.

Finally, Fényes assumes a Lagrangian function and by using a variation principle deduces the Fokker equation and the Schrödinger equation for a certain class of systems.

This paper is an attempt to show that Fényes' theory is unsatisfactory. Many points of the theory which seem open to debate will not be discussed here because they are part of long verbal arguments which would require considerable analysis. Some further discussion may be found in the writer's thesis (Nicholson 1953). The theory has also been criticized by Weizel (1953).

II. FÉNYES' MARKOFF RELATION AND QUANTUM MECHANICS

The Lagrangian function used by Fényes appears to be of arbitrary form. Fényes has not shown that the resulting Schrödinger equation and its physically admissible solutions are equivalent to his original equations and their allowable solutions.

We shall now show that quantum mechanics cannot be based solely on Fényes' original integral relations which define a Markoff process, but that further restrictions are necessary on the class of solutions of the Markoff equations before these solutions can be taken to describe quantum-mechanical motions.

Let us consider a quantum-mechanical system for which the Schrödinger equation has only a discrete set of energy eigenvalues E_i and associated physically-admissible normalized linearly-independent eigenfunctions ψ_i . We suppose that no continuous ranges of eigenvalues exist and that none of the energy eigenstates is degenerate. One such system is the one-dimensional linear harmonic oscillator which has a denumerably infinite set of discrete energy eigenstates. Let ψ_1, ψ_2 be any two of the energy eigenfunctions. Then $\psi_1^* \psi_1, \psi_2^* \psi_2$ are time independent and are interpreted as probability densities. If

Fényes' theory is applicable then the motion of the system can be represented as a Markoff process defined by the relations

$$w(y,t) = \int_{-\infty}^{\infty} v(x,s; y,t) w(x,s) dx, \quad \dots\dots\dots (1)$$

$$\int_{-\infty}^{\infty} w(y,t) dy = 1, \quad \dots\dots\dots (2)$$

$$w(y,t) = \psi^*(y,t) \psi(y,t), \quad \dots\dots\dots (3)$$

with obvious meanings. Fényes supposes that any non-negative solution of (1) and (2) represents a possible mode of motion of the system. If Fényes' theory is applicable we must suppose that the time-independent distributions

$$w_1(y) = \psi_1^* \psi_1, \quad w_2(y) = \psi_2^* \psi_2$$

are solutions of (1) and (2). But it is then clear that any linear combination

$$w(y) = kw_1(y) + (1-k)w_2(y), \quad k \text{ constant, } 0 \leq k \leq 1,$$

also satisfies (1) and (2) and is non-negative and so must represent a possible stationary state of the system.

Since k can assume a continuous infinity of values in the range (0,1), Fényes' theory predicts an infinite set of power continuum of stationary states of the system. But the system considered has only a discrete set of energy eigenvalues, so that the set of stationary states is at most denumerably infinite. Hence Fényes' equations permit an infinitely greater number of stationary states than are allowed by quantum mechanics, for certain systems. Such a continuum of states can be constructed from each pair of discrete states.

Hence any theory, like that of Fényes, based on the equations (1), (2), and (3) cannot be equivalent to quantum mechanics, which is based on a linear equation for the probability amplitude ψ . This conclusion is also suggested by a discussion of the principle of superposition of states given by Feynman (1948). The solutions w of (1) and (2) would need to satisfy some further equation involving w non-linearly before the resulting theory could be equivalent to quantum mechanics.

In a representation of quantum mechanics as a form of statistical mechanics by Moyal (1949), the state of the system is described by a phase-space distribution function $F(p,q;t)$ which satisfies an equation of Markoff form, namely,

$$F(p,q;t) = \iint K(p,q \mid p_0,q_0;t-t_0) F(p_0,q_0;t_0) dp_0 dq_0.$$

As for Fényes' theory, we can see that for certain systems the possible normalized time-independent solutions of such an equation, if they exist at all, are more numerous than the stationary states of the systems as given by quantum mechanics, so that quantum mechanics cannot be based solely on equations of this form. However, Moyal defines his F directly in terms of the Schrödinger amplitude ψ , and, although F satisfies the equation above, this equation does not define the class of admissible F in his theory.

III. GENERALITY OF FÉNYES' THEORY

The Schrödinger equation deduced by Fényes applies only to a certain class of systems, in which the interactions between the particles and the external field can be expressed completely in terms of scalar potential functions. The Schrödinger equation given by Fényes is not applicable, for instance, to a particle moving in an external electromagnetic field.

Fényes' form of the Schrödinger equation is in position-coordinate language. If his basic variables were declared to be, say, linear momenta instead of position coordinates, then the Schrödinger equation would be in momentum language, but the class of known physical systems to which it would then apply would be very small: for the Hamiltonian would have to involve the position coordinates q_i only in the form of the term $\sum_i q_i^2$. The harmonic oscillator is a member of this small class. Further, if Fényes' basic variables were linear momenta, then the various "velocities" and the "uncertainty relation" which occur in his theory would have no familiar interpretation.

It appears to be difficult to extend Fényes' method to yield the wave equation in coordinate language for more general systems, or to deduce a satisfactory eigenvalue-eigenfunction equation in other languages, or for other dynamical variables.

IV. FÉNYES' ψ AND THE PROBABILITY AMPLITUDE OF QUANTUM MECHANICS

We now show that it is impossible to identify Fényes' ψ with the probability amplitude ψ of quantum mechanics.

Consider two energy eigenstates ψ_1 , ψ_2 of a suitable quantum-mechanical system, which eigenstates belong to the discrete unequal eigenvalues E_1 , E_2 . Then we can write

$$\psi_1(q,t) = u_1(q)e^{-2\pi i E_1 t/\hbar}, \quad \psi_2(q,t) = u_2(q)e^{-2\pi i E_2 t/\hbar}.$$

In the quantum-mechanical case, Fényes' definition of ψ becomes

$$\psi = \alpha e^{2\pi i(\sigma + i\Sigma)/\hbar},$$

where α can be taken as a real positive constant. Equating the probability densities in the two theories we obtain

$$\psi_1^* \psi_1 = u_1^* u_1 = \alpha_1^2 e^{-4\pi \sigma_1/\hbar}$$

and similarly for ψ_2 , so that

$$-\frac{2\pi\sigma_1}{\hbar} = \ln \frac{|u_1|}{\alpha_1}, \quad -\frac{2\pi\sigma_2}{\hbar} = \ln \frac{|u_2|}{\alpha_2}.$$

Allowing for an arbitrary phase factor, we equate the probability amplitudes from the two theories to obtain

$$\psi_1 = \alpha_1 e^{i \ln(|u_1|/\alpha_1) - 2\pi i \Sigma_1/\hbar} = |u_1| e^{-2\pi i \Sigma_1/\hbar} = u_1 e^{-2\pi i \eta_1/\hbar} e^{-2\pi i E_1 t/\hbar},$$

and similarly for ψ_2 , where η_1 is a constant (cf. Witmer and Vinti 1935).

We can write

$$u_1 = |u_1| e^{-i\delta_1(q)},$$

where $\delta_1(q)$ is a real function of q . Then

$$\Sigma_1 = \delta_1 + E_1 t + n_1 \hbar + \eta_1, \quad n_1 \text{ an integer.}$$

Similarly

$$\Sigma_2 = \delta_2 + E_2 t + n_2 \hbar + \eta_2.$$

But by Fényes' definitions

$$\Sigma = s + \sigma,$$

where $s = s(q, t)$ is independent of the state of the system and depends only upon its nature. Therefore

$$\sigma_1 + s = \delta_1 + E_1 t + n_1 \hbar + \eta_1,$$

$$\sigma_2 + s = \delta_2 + E_2 t + n_2 \hbar + \eta_2,$$

and

$$\sigma_1 - \sigma_2 - \delta_1 + \delta_2 = (E_1 - E_2)t + \hbar(n_1 - n_2) + (\eta_1 - \eta_2).$$

But this is impossible, as the right side is a non-zero function of t (since $E_1 \neq E_2$) but not of q , whereas the left side is a function of q only. Hence we cannot equate the probability amplitudes ψ in the two theories, and Fényes' theory is not a possible representation of quantum mechanics.

V. TOTAL STOCHASTIC VELOCITY AND THE UNCERTAINTY RELATION

We next discuss Fényes' total stochastic velocity of components c_i , and his uncertainty relation. If $w(y, t)$ is Fényes' space-time distribution function then

$$c_i = a_i - \frac{1}{w} \sum_k \frac{\partial}{\partial y_k} (b_{ik} w), \quad \dots \dots \dots (4)$$

where $a_i(y, t)$ and $b_{ik}(y, t)$ depend upon the nature of the system but not upon its state. The important thing here is that c_i is a precise function of position and time.

Fényes defines the means \bar{c}_i, \bar{y}_i and the variances $\Delta c_i, \Delta y_i$ of the stochastic velocity components c_i and the configuration coordinates y_i with respect to w over all configuration space, and deduces a lower bound for the product $\Delta y_i \Delta c_i$. A similar relation has been given by Fürth (1933) for the case of one-dimensional Brownian motion of a particle. Fényes attempts to identify this uncertainty relation with the Heisenberg uncertainty relation of quantum mechanics. To do this for a particle it is necessary to identify c_i with p_i/m where m is the particle mass and p_i is the quantum-mechanical linear momentum component corresponding to the coordinate y_i . But c_i is a function of the configuration coordinates y_i as well as the time, whereas the quantum-mechanical momentum distribution is a function of the p_i and time only, for any state of the system. Any dependence of p_i on the configuration coordinate y_i would amount to a violation of the Heisenberg uncertainty principle. Hence it is impossible to identify c_i with p_i/m or Δc_i with $\Delta p_i/m$. The "uncertainties" Δc_i and Δp_i have different meanings.

This basic difference is reflected in Fényes' definition of a stochastic velocity operator c_i which satisfies

$$c_i w = c_i w$$

for any admissible state w of a system, that is, every possible state of the system is an eigenstate of c_i . In quantum mechanics only a subset of all admissible states ψ of a system are eigenstates of p_i , that is, satisfy

$$p_i \psi = p_i \psi.$$

Later in his paper Fényes modifies his velocity definition to define a velocity of components c'_i which, he states, agree with the quantum-mechanical velocity components. But again c'_i is a precise function of position and time, and thus cannot be identified with p_i/m . Further c'_i is in general complex, whereas p_i is always real.

VI. CAUSAL AND STATISTICAL DESCRIPTION

Some of the mathematical arguments in Fényes' paper appear doubtful. We shall only consider one such argument here, namely, Fényes' proof that the addition of further parameters y_{n+1}, \dots, y_N to the basic parameters y_1, \dots, y_n cannot cause the scattertensor b_{ik} to vanish in the completed parameter system if it does not vanish in the original system. This proposition is probably correct though the argument seems wrong. From this proposition Fényes concludes that a causal description cannot be obtained from a statistical description by the addition of further parameters.

In the original parameter system the transition probability density is denoted by $v(y_1, \dots, y_n, t, z_1, \dots, z_n, t + \Delta)$ while in the completed set it is $V(y_1, \dots, y_N, t, z_1, \dots, z_N, t + \Delta)$.

The scattertensors in the two systems are

$$\left. \begin{aligned} b_{ik}(y_1, \dots, y_n, t) &= \lim_{\Delta \rightarrow 0} \frac{1}{2\Delta} \int (z_i - y_i)(z_k - y_k) v dz_1 \dots dz_n, \\ \beta_{ik}(y_1, \dots, y_N, t) &= \lim_{\Delta \rightarrow 0} \frac{1}{2\Delta} \int (z_i - y_i)(z_k - y_k) V dz_1 \dots dz_N. \end{aligned} \right\} \dots (5)$$

Fényes states that it is obvious that

$$v = \int V dz_{n+1} \dots dz_N. \dots \dots \dots (6)$$

But V is a function of y_{n+1}, \dots, y_N while v is not. There is a class of functions V for which $\int V dz_{n+1} \dots dz_N$ is independent of y_{n+1}, \dots, y_N , but Fényes has not shown that the functions V considered in his paper belong to this class.

Fényes' conclusion can be drawn, however, on the basis of a plausible assumption. It seems clear from the meaning of a probability density that

$$w(y_1, \dots, y_n, t) = \int W(y_1, \dots, y_N, t) dy_{n+1} \dots dy_N,$$

where w and W are the probability densities in the two systems. Then equation (1) applied in the two systems yields

$$\begin{aligned} & \int w(y_1, \dots, y_n, t) v(y_1, \dots, y_n, t, z_1, \dots, z_n, t + \Delta) dy_1 \dots dy_n \\ &= \iint W(y_1, \dots, y_N, t) V(y_1, \dots, y_N, t, z_1, \dots, z_N, t + \Delta) dy_1 \dots dy_N dz_{n+1} \dots dz_N, \end{aligned}$$

which suggests that

$$wv = \iint W V dy_{n+1} \dots dy_N dz_{n+1} \dots dz_N \dots \dots \dots (7)$$

is true, instead of (6).

In Fényes' theory the probability of a transition from the point (y_1, \dots, y_n) at t to the region $dz_1 \dots dz_n$ about (z_1, \dots, z_n) at $t + \Delta$ is

$$v dz_1 \dots dz_n,$$

and similarly for V . Classically, the probability of transition from (y_1, \dots, y_n) at t to $dz_1 \dots dz_n$ about (z_1, \dots, z_n) at $t + \Delta$ is

$$dz_1 \dots dz_n \int V dz_{n+1} \dots dz_N,$$

and the probability of transition from (y_1, \dots, y_n) at t to $dz_1 \dots dz_n$ about (z_1, \dots, z_n) must then be

$$\begin{aligned} & dz_1 \dots dz_n \int \left\{ \text{Relative probability of } (y_{n+1}, \dots, y_N) \text{ at } t \right. \\ & \quad \times \left. \int V dz_{n+1} \dots dz_N \right\} dy_{n+1} \dots dy_N \\ &= dz_1 \dots dz_n \int \left\{ \frac{W(y_1, \dots, y_N, t)}{\int W dy_{n+1} \dots dy_N} \cdot \int V dz_{n+1} \dots dz_N \right\} dy_{n+1} \dots dy_N \\ &= \frac{dz_1 \dots dz_n}{w(y_1, \dots, y_n, t)} \iint W V dy_{n+1} \dots dy_N dz_{n+1} \dots dz_N. \end{aligned}$$

But this transition probability is $v dz_1 \dots dz_n$. Hence (7) is true for a classical theory, and we assume it to hold true in Fényes' theory as it is consistent with Fényes' relations between v and w , which are of a classical kind.

We now consider the integral

$$I = \frac{1}{2\Delta} \int (z_i - y_i)(z_k - y_k) V(y, t, z, t + \Delta) W(y, t) dy_{n+1} \dots dy_N dz_1 \dots dz_N,$$

for $i \leq n, k \leq n$.

Clearly,

$$\begin{aligned} I &= \frac{1}{2\Delta} \int (z_i - y_i)(z_k - y_k) \left\{ \int V W dy_{n+1} \dots dy_N dz_{n+1} \dots dz_N \right\} dz_1 \dots dz_n \\ &= \frac{1}{2\Delta} \int (z_i - y_i)(z_k - y_k) w(y, t) v(y, t, z, t + \Delta) dz_1 \dots dz_n, \text{ by equation (7),} \\ &= w(y, t) \frac{1}{2\Delta} \int (z_i - y_i)(z_k - y_k) v(y, t, z, t + \Delta) dz_1 \dots dz_n. \end{aligned}$$

But also

$$I = \int W(y, t) \left\{ \frac{1}{2\Delta} \int (z_i - y_i)(z_k - y_k) V dz_1 \dots dz_N \right\} dy_{n+1} \dots dy_N.$$

We equate these expressions for I , and let $\Delta \rightarrow 0$. Then the definitions (5) yield, for $i, k \leq n$,

$$w(y_1, \dots, y_n, t) b_{ik}(y_1, \dots, y_n, t) = \int W(y_1, \dots, y_N, t) \beta_{ik}(y_1, \dots, y_N, t) dy_{n+1} \dots dy_N.$$

Hence, if, for $i, k \leq n$,

$$b_{ik} \neq 0,$$

then

$$\beta_{ik} \neq 0.$$

This is Fényes' conclusion, and shows that the addition of further coordinates y_{n+1}, \dots, y_N cannot cause the scattertensor to vanish in the completed system if it does not vanish in the original system. Fényes also concludes that, for $i, k \leq n$, $\beta_{ik} = b_{ik}$, but this does not follow from the present argument, and is doubtful since β_{ik} is in general a function of y_{n+1}, \dots, y_N while b_{ik} is not.

Fényes next proves that a system can only be assigned a phase-space density function $f(u, y, t)$ and be treated by the methods of statistical mechanics if the scattertensor b_{ik} vanishes identically. Some objections to this proof are stated in the writer's thesis (Nicholson 1953).

VII. ACKNOWLEDGMENT

The writer desires to express his gratitude to Dr. R. Fürth for his kindness and encouragement during the course of this work.

VIII. REFERENCES

- FÉNYES, I. (1952).—*Z. Phys.* **132**: 81.
 FEYNMAN, R. P. (1948).—*Rev. Mod. Phys.* **20**: 368-9.
 FÜRTH, R. (1933).—*Z. Phys.* **81**: 143.
 KOLMOGOROFF, A. N. (1931).—*Math. Ann.* **104**: 415.
 KOLMOGOROFF, A. N. (1933).—*Math. Ann.* **108**: 149.
 MOYAL, J. E. (1949).—*Proc. Camb. Phil. Soc.* **45**: 99.
 VON NEUMANN, J. (1932).—"Mathematische Grundlagen der Quantenmechanik." Ch. 3, Section 2; Ch. 4, Sections 1 and 2. (Springer: Berlin.)
 NICHOLSON, A. F. (1953).—M.Sc. Thesis, University of London.
 TOLMAN, R. C. (1938).—"The Principles of Statistical Mechanics." p. 219. (Oxford Univ. Press.)
 WEIZEL, W. (1953).—*Z. Phys.* **134**: 264.
 WITMER, E. E., and VINTI, J. P. (1935).—*Phys. Rev.* **47**: 538.

SOME ELECTRON COLLISION CROSS SECTIONS OF CaII

By J. T. JEFFERIES*

[Manuscript received September 29, 1953]

Summary

The cross sections for electron collision excitation of the transitions $4S_{1/2}-4P_{1/2}$ and $4P_{3/2}$ in CaII have been calculated using a distorted wave Born approximation. For electrons of energy 4 eV, the calculated values are respectively $6.6\pi a_0^2$ and $9.9\pi a_0^2$.

I. INTRODUCTION

The interpretation of the physical properties of the solar chromosphere and of various phenomena connected with it is assisted by calculations of radiation emitted by model atmospheres, such, for example, as have been performed for hydrogen (Giovanelli 1949; Jefferies 1953). To extend these calculations to the important case of CaII the rates of collisional excitation are needed, particularly for transitions between the $4S$ and $4P$ levels, which give rise to the H and K lines. These cross sections are of interest also, as shown by Miyamoto (1953), in assessing the role played by non-coherent scattering in the formation of these lines. In the following, results are given of a calculation of the cross sections for electron excitation of the two transitions $4s^2S_{1/2} \rightarrow 4p^2P_{1/2}$ and $4p^2P_{3/2}$.

II. THE COLLISION CROSS SECTIONS

In atomic units the cross section for excitation, by electron collision, of the transition $n \rightarrow n'$ may be written, see, for example, Mott and Massey (1949),

$$\sigma(n \rightarrow n') = \frac{1}{4\pi^2} \frac{k'}{k} \frac{1}{2(2J+1)} \sum_{m_s, m_{s'}, M_J, M_{J'}} \int |(n|V|n')|^2 d\Omega, \quad \dots \quad (1)$$

k and k' being the momenta of the incident and scattered electrons, m_s and $m_{s'}$ their spin components, M_J and $M_{J'}$ the components of the total angular momentum of the initial and final states of the ion, and $d\Omega$ the solid angle into which the electron is scattered. The symbols n and n' involve both the ion and colliding electron. The factor $2(2J+1)$ —equal to 4 in the present case—arises from an averaging over the two possible spin orientations of the incident electron and the $2J+1$ values of M_J .

The term $(n|V|n')$ is defined by

$$(n|V|n') = \int \Psi_n^* V \Psi_{n'} d\tau,$$

* Division of Physics, C.S.I.R.O., University Grounds, Sydney.

V being the interaction energy, and the Ψ 's representing wave functions of the complete system, ion and incident electron. The Ψ 's should strictly be anti-symmetrical combinations in the coordinates (space and spin) of the colliding and ionic electrons. However, for transitions of interest here, it appears (Bates *et al.* 1950) that a better approximation is obtained if exchange effects are neglected, and we accordingly take Ψ to be a simple product of wave functions,

$$\Psi_n = \psi_A(1)\chi_k(2), \quad \Psi_{n'}^* = \psi_B^*(1)\chi_{k'}^*(2),$$

where the ionic and colliding electrons are designated respectively by the numbers 1 and 2.

The $\psi_A(1)$ and $\psi_B(1)$ are taken as appropriate linear combinations—as given for example by Condon and Shortley (1935)—of one-electron wave functions, whose radial components are tabulated by Hartree and Hartree (1935), the angular components being of the central field type.

For the incident electron in the field of the ion, we find

$$\chi_k = r^{-1}k^{-1/2} \sum_{l=0}^{\infty} L_l(k, r) P_l(\cos \theta) \begin{Bmatrix} \alpha \\ \beta \end{Bmatrix}, \quad \dots \quad (2)$$

where

$$L_l(k, r) = (2l+1)! e^{i\eta_l} G_l(r),$$

$G_l(r)$ being the solution of the differential equation

$$G_l''(r) + [k^2 + 2Z_p/r - l(l+1)/r^2] G_l(r) = 0, \quad \dots \quad (3)$$

which goes to zero as r^{l+1} and asymptotes to $k^{-1/2} \sin(kr + \eta_l)$, η_l being a phase whose magnitude is not required for our problem. The symbols α and β represent spin functions and $P_l(\cos \theta)$ is the Legendre polynomial. The term $2Z_p/r$ in equation (3) represents the potential of the free electron in the CaII field and can be found from results given by Hartree and Hartree (1935) who tabulate $2Z_p$ for CaI and CaIII. For CaII the mean of these has been used. The wave function of the scattered electron may be written,

$$\chi_{k'} = k'^{-1/2} r^{-1} \sum_{l'=0}^{\infty} L_{l'}(k', r) P_{l'}(\cos \Theta) \begin{Bmatrix} \alpha \\ \beta \end{Bmatrix}, \quad \dots \quad (4)$$

Θ being the angle between the momentum vector of the scattered electron and the radius vector.

Taking the interaction energy V to be $1/r_2 - 1/r_{12}$ and using well-known expansions in terms of associated Legendre polynomials, we may now calculate $(n|V|n')$ and hence the cross section. The results obtained, for $k^2 = 0.30$, are

$$\sigma(4s^2S_{1/2} - 4p^2P_{1/2}) = 6.6\pi a_0^2,$$

$$\sigma(4s^2S_{1/2} - 4p^2P_{3/2}) = 9.9\pi a_0^2.$$

where a_0 is the radius of the first Bohr orbit of hydrogen.

In obtaining these, a sum over the index l of equation (2) has been made. It may be shown (Bohr, Peierls, and Placzek 1949) that, for any value of l , the contribution to the cross section can be no greater than $(2l+1)\pi a_0^2/k^2$. The values obtained here were checked against this condition at each stage, and were found to be compatible with it in all cases; in only one ($l=1$) did it approach the limiting value. In fact, the results given above are effectively those due to the partial wave with $l=1$.

III. ACKNOWLEDGMENTS

The author's thanks are due to Mr. E. R. Hill of the Division of Radiophysics, C.S.I.R.O., for valuable advice, to Miss J. Y. Ward for performing most of the computations, and to Dr. R. G. Giovanelli for constant encouragement.

IV. REFERENCES

- BATES, D. R., FUNDAMINSKY, A., LEECH, J. W., and MASSEY, H. S. W. (1950).—*Phil. Trans*, A **243**: 93.
- BOHR, N. F., PEIERLS, R., and PLACZEK, G. (1949).—Quoted by Mott and Massey (1949).
- CONDON, E. U., and SHORTLEY, G. H. (1935).—"The Theory of Atomic Spectra." (Cambridge Univ. Press.)
- GIOVANELLI, R. G. (1949).—*Mon. Not. R. Astr. Soc.* **109**: 298.
- HARTREE, D. R., and HARTREE, W. (1935).—*Proc. Roy. Soc. A* **149**: 210.
- JEFFERIES, J. T. (1953).—*Aust. J. Phys.* **6**: 22.
- MIYAMOTO, S. (1953).—*Z. Astrophys.* **31**: 282.
- MOTT, N. F., and MASSEY, H. S. W. (1949).—"The Theory of Atomic Collisions." 2nd Ed. (Oxford Univ. Press.)

THE CONTINUOUS RADIATIVE ABSORPTION CROSS SECTION OF FeXIV AND THE CORONAL TEMPERATURE

By ALMA WERNER*

[Manuscript received September 17, 1953]

Summary

The continuous radiative absorption cross section of FeXIV was calculated, using a Hartree wave function to evaluate the matrix element for recombination to the ground state. This matrix element was considerably smaller than the value obtained by Hill (1950, 1951) using the hydrogen-like approximation, but the total cross section for recombination was not greatly different. Balancing the rates of recombination and collision ionization in the solar corona then gave a temperature of about 2×10^6 °K, compared with Hill's value of 1×10^6 °K.

I. INTRODUCTION

FeXIV is one of the most important of the highly ionized atoms in the solar corona (Woolley and Allen 1948) and, by balancing the rates of radiative recombination and collision ionization of this ion, one may deduce the electron agitation temperature of the corona. Estimates of these rates were made by Woolley (1947) using classical methods for the ionization cross section and taking a value calculated for OI at 6000 °K, modified by assuming a certain dependence of the cross section on the atomic number and temperature, for the radiative recombination cross section. Improved estimates have since been made by Hill (1951) using wave mechanical methods. Both calculations gave a coronal temperature of close to 10^6 °K. Independent methods of estimating this temperature have given somewhat higher values (see Woolley 1947).

Hill's calculation of the recombination coefficient of FeXIV was made using an appropriate modification of Wessel's formula for the recombination of an electron with a proton. Detailed calculations of recombination coefficients using accurate atomic wave functions have now been made for several neutral atoms and singly ionized atoms, and it appears that the hydrogen-like approximation can sometimes give results seriously in error, as a result of considerable cancellation in the integral for the matrix element. As it seemed likely that such sensitivity of the calculation to the form of the radial wave function would occur, *a fortiori*, for highly charged ions, and as a Hartree field for FeXIV has become available since Hill's paper was written (Gold 1949) an accurate calculation of the recombination coefficient of FeXIV was undertaken.

* Physics Department, University of Melbourne.

II. RADIATIVE RECOMBINATION TO THE GROUND STATE

Quantum theory gives the following expression for the total cross section of an atomic system for radiative recombination

$$Q_R^t(k^2) = \sum_s Q_R^s(k^2) = \frac{128\pi^5 m e^2}{3\hbar^2 c^3 k} \sum_s \nu_{ks}^3 \left| \int \psi_s^* \mathbf{r} \psi_k d\tau \right|^2, \quad \dots \quad (1)$$

ψ_s being the wave function for the bound state, with energy E_s , into which the electron falls, and ψ_k the continuous wave function for the incident electron with momentum $\hbar k/2\pi$ before capture. The frequency of the emitted radiation is ν_{ks} .

If $V(r)$ is the potential of the ion field to which the free electron is captured, then

$$\psi = (2l+1)! e^{i\eta_l} k^{-\frac{1}{2}} r^{-1} \sum_{l=0}^{\infty} G_l(r) P_l(\cos \theta), \quad \dots \quad (2)$$

where G_l is the solution of the differential equation

$$G_l'' + [k^2 - 2V(r) - l(l+1)r^{-2}]G_l = 0, \quad \dots \quad (3)$$

which vanishes at the origin and has the asymptotic form $k^{-\frac{1}{2}} \sin(kr + \eta_l)$, η_l being a phase shift whose value we shall not require.

The matrix elements sensitive to the form of the wave functions will be those for capture to the ground state of the ion, namely, the state $n=3$, $l=1$. The hydrogen-like approximation used by Hill will be sufficiently accurate for capture to excited states. From the usual selection rule, capture to the ground ($3p$) state can only take place from an s or d state in the continuum, involving the functions G_0 and G_2 respectively. G_0 is the only one of the G 's which differs markedly from a Coulomb wave function for $Z=16.4$ (corresponding to an effective nuclear charge of $16.4e$ in the vicinity of the $3p$ electron), since the $l(l+1)r^{-2}$ term in (3) swamps any deviation of $V(r)$ from the Coulomb field for small r , and at much larger distances the Coulomb field is small. Only $Q_R^{3,1}$, therefore, was calculated using formula (1) with accurate wave functions.

The Hartree field and ground state wave functions have been obtained for FeXIV by Gold (1949), and the required values of $V(r)$ to be used in (3) were found by subtracting the potential $\varphi(r)$ of the $3p$ electron from the Hartree field of FeXIV, where

$$\varphi(r) = \frac{1}{r} \int_0^r \psi^2 d\xi + \int_r^\infty \xi^{-1} \psi^2 d\xi, \quad \dots \quad (4)$$

$\psi(\xi)$ being the (radial) Hartree wave function (averaged over all angles) for the $3p$ electron. G_0 was then found by numerical integration of (3).

The form of G_0 and ψ_s were found to be such as to cause considerable cancellation in the integrand in (1), and the values obtained for $Q_R^{3,1}$ by adding together the contribution from the s and d states of the continuum were several times smaller than the corresponding values obtained by Hill, as will be seen by comparing the second and third columns in Table 1.

III. RADIATIVE RECOMBINATION TO THE EXCITED STATES

The hydrogen-like approximation, as used by Hill, will be sufficiently accurate for capture to the excited states. Hill (1950) has given graphs for various Q_R^n , n being the total quantum number, based on matrix elements calculated by Wessel for $n=1, 2, 3$ and by Bates for not too low energies. At low energies, for $n>3$ and for $(ka)^{-1}>20$ (where $k=2\pi mv/h$ and $a=a_0/Z$), one uses Kramer's classical quantum theory expression

$$Q_R^n = \frac{16e^2h}{3^{3/2}c^3m^2} (ka)^{-4} \frac{1}{n\{n^2 + (ka)^{-2}\}^2} \dots\dots\dots (5)$$

When allowance is made for the levels which are already filled, we have for the recombination cross section of FeXIV

$$Q_R = Q_R^t - (Q_R^{1,0} + Q_R^{2,0} + Q_R^{2,1} + Q_R^{3,0}). \dots\dots\dots (6)$$

Values of the various Q_R^n on the right side of (6) were obtained from Hill's graphs, and the necessary correction made from the more accurate values of $Q_R^{3,1}$ found as above. The final value so obtained for Q_R is shown in the last column of Table 1. It is seen that the major contribution to Q_R comes from capture to the higher states, so that Hill's considerable over-estimation of the cross section for capture to the ground state does not have too serious consequences.

TABLE 1
RECOMBINATION CROSS SECTIONS

$(ka)^{-1}$	$Q_R^{3,1}$ Werner (10^{-22} cm 2)	$Q_R^{3,1}$ Hill (10^{-22} cm 2)	Q_R^t Hill (10^{-22} cm 2)	Q_R (10^{-22} cm 2)
4	0.07	3.1	55	13
5	0.13	5.4	97	30
6	0.20	8.2	154	55
10	0.73	24.6	537	249
14	1.8	49	1200	635
20	5.1	100	2720	1550
25	9.8	157	4580	2750

IV. TEMPERATURE OF THE SOLAR CORONA

If the peak of the Maxwell distribution came near the plateau of the ionization curve, one might reasonably estimate the temperature of the solar corona by finding the particular energy for which $Q_R=Q_I$. But the peak falls far below the threshold for ionization (below 100 eV compared with a threshold of 376 eV). The temperature of the corona must therefore be obtained by finding the temperature for which $(Q_R)_{Av.}=(Q_I)_{Av.}$, where the suffix Av. indicates that the values of Q_R and Q_I must be averaged over a Maxwellian distribution of electron energies.

Using Table 1, $(Q_R)_{Av.}$ may be obtained for a few temperatures, and, using Hill's ionization curves, $(Q_I)_{Av.}$ may be obtained for the same temperatures.

The point of intersection of the curves of $(Q_R)_{Av.}$ and $(Q_I)_{Av.}$ against temperature gives the required temperature T of the corona.

It is found that about 0.4 of the total value of $(Q_R)_{Av.}$ comes from the low energy tail of the Maxwell distribution below an energy of 4 eV corresponding to $ka < 1/30$. This is due to the rapid rise in the Q_R 's towards infinite values as ka tends to zero, so that the value of $Q_R \cdot dN/dE$, where dN is the number of electrons with energies between E and $E+dE$, when plotted as a function of E , is found to rise again as E falls to low energies. Hill appears to have avoided this embarrassing behaviour by applying a "cut-off" below 4 eV. However, the integrated value of Q_R is found to be finite, and its value over the range $ka=0$ to $1/30$ was obtained by replacing the summation over n in (5) by an integration over n , adjusting the lower limit suitably, and finally integrating over ka . Owing to the high values of Q_R at low energies, a high radiative absorption of electrons in this range will occur, and one might expect an appreciable deviation in the energy distribution from the Maxwell values at low energies, especially as most of the electron-electron collisions will be small-angle, resulting in a slow interchange of electron energies. On the other hand, this effect will be damped out to some extent by the overwhelming predominance of elements other than iron in the solar corona. Such a deviation could exert a considerable change in the value obtained by integrating Q_R .

The values of the averaged Q 's obtained are given in Table 2.

TABLE 2
AVERAGE CROSS SECTIONS

T (10^6 °K)	1.0	1.5	2.0	3.0
$(Q_R)_{Av.}$ (10^{-20} cm ²)	3.1		1.4	1.1
$(Q_I)_{Av.}$ (a) (10^{-20} cm ²)		0.66	1.27	2.43
(b) (10^{-20} cm ²)		0.53	1.00	1.90

There is some uncertainty in the values of Q_I obtainable from Figure 2 of Hill's paper. The values (a) in Table 2 were obtained using the mean of curves 1 and 2, and the values (b) using curve 4. These gave the following values for T :

Values (a): 2.1×10^6 °K,

Values (b): 2.25×10^6 °K.

These two values are to be compared with the value 1.1×10^6 °K obtained by Hill, and with the values 2.3×10^6 , 3.3×10^6 , and $2.5-6.5 \times 10^6$ °K at different heights above the limb, obtained by entirely different methods (see Woolley 1947).

However, the experimental data giving FeXIV as the most abundant of the various ionized states of iron cannot be assumed to apply to the quiescent corona, but will, in fact, be bound to include a proportion (unknown) of coronal "hot spots". Similarly, the other values of the coronal temperature, obtained by various other methods, would also be some kind of mean between hot spot

and quiescent temperatures. The value of the temperature for the quiescent corona, given by the minimum solar radio noise, is about 0.5×10^6 °K, which is considerably below the above-mentioned temperatures.

V. ACKNOWLEDGMENT

The author is indebted to Associate Professor C. B. O. Mohr for suggesting this problem and for his continued interest.

VI. REFERENCES

- GOLD, M. T. (1949).—*Mon. Not. R. Astr. Soc.* **109**: 471.
HILL, E. R. (1950).—M.Sc. Thesis, University of Melbourne.
HILL, E. R. (1951).—*Aust. J. Sci. Res. A* **4**: 437.
WOOLLEY, R. v. D. R. (1947).—*Aust. J. Sci.* **10** (2): Suppl.
WOOLLEY, R. v. D. R., and ALLEN, C. W. (1948).—*Mon. Not. R. Astr. Soc.* **108**: 293.

SPACE CHARGE WAVE AMPLIFICATION IN A SHOCK FRONT AND THE FINE STRUCTURE OF SOLAR RADIO NOISE

By HARI K. SEN*

[*Manuscript received August 17, 1953*]

Summary

The Mott-Smith (1951) interpolation method gives a non-Maxwellian velocity distribution for the particles in a shock front. The dispersion equation corresponding to the non-Maxwellian distribution is derived by Vlasov's (1945) formula. The roots of the dispersion equation indicate frequency bandwidths of space charge wave amplification that decrease with the shock strength. It is suggested, in agreement with Denisse and Rocard (1951), that the storm bursts of narrow bandwidth originating in shock fronts constitute the elementary fine-structure components of solar radio noise bursts.

I. INTRODUCTION

Solar radio noise bursts have a fine structure (Blum and Denisse 1950 ; Reber, unpublished data ; Wild 1951) in the frequency bandwidth of amplification, of the order of a few per cent. of the base frequency. Denisse and Rocard (1951) suggested the origin of these elementary storms to be the space charge wave amplification in a shock front. They drew attention to the fact that the electron velocity distribution in the shock front of an ionized gas will depart from the Maxwellian one, developing a secondary hump at a velocity exceeding the mean thermal motion of the electrons. Applying the Landau criterion (Landau 1946 ; Bohm and Gross 1949), they concluded that the secondary hump would be responsible for the excitation of electronic oscillations and indicated its application to the observed fine structure of solar radio noise bursts.

Denisse and Rocard, however, used the Enskog-Chapman approximation (Chapman and Cowling 1939), which, on account of its slow convergence, is applicable only to very weak shocks, as pointed out by Wang Chang (1948). They admitted in their paper that, though there was every reason to believe that the effects discussed by them would augment with the strength of the shock, their analysis did need extension to strong shocks.

In a recent paper, Mott-Smith (1951) indicated a promising method of approach to the velocity distribution in a strong shock. He pointed out that the distribution of molecular velocities in a strong shock wave in a gas would be bimodal in view of the fact "that a considerable number of the Maxwellian molecules of the bounding supersonic and subsonic streams penetrate into the centre of the shock". He found the velocity distribution function on the

* National Bureau of Standards, Washington 25, D.C., U.S.A.

assumption that it would be a sum of two Maxwellian terms with temperatures and mean velocities corresponding to the subsonic and supersonic streams. He further showed that his distribution function was an approximate stationary solution of the Boltzmann equation for strong shocks.

II. THE DISPERSION EQUATION

We have followed in this paper* the Mott-Smith approach, as we believe that it is applicable to strong shocks and also brings out clearly the physical concept of the bimodal velocity distribution.

We assume the gas to be completely ionized with equal numbers of positive ions and electrons. We also neglect the small polarization effect (Denisse and Rocard 1951) and assume the electrons and ions to have a common mass velocity. On account of our neglect of the polarization effect, the concentration functions will refer indifferently to the ion or the electron.† The Mach number, however, will refer to the ions, which mainly determine the shock wave.

We shall use the following notation :

- e , charge of electron (in e.s.u.),
- m , mass of the electron (or ion),
- κ , Boltzmann's constant,
- u_α , stream velocity before shock,
- u_β , stream velocity behind shock,
- T_α , stream temperature before shock,
- T_β , stream temperature behind shock,
- γ , ratio of specific heats,
- $a = 2\gamma/(\gamma - 1)$,
- M , Mach number of stream before shock $= u_\alpha \sqrt{\{(a-2)/a\}(m/\kappa T_\alpha)}$,
- n_0 , particle concentration before shock,
- n_0' , particle concentration behind shock,
- n , particle concentration in shock centre,
- u , particle velocity along the direction of propagation of shock,
- \bar{u} , common mass velocity of electrons and ions,
- $U = u - \bar{u}$,
- $U' = \sqrt{(m/2\kappa T_\alpha)} U$,
- k = wave number of A.C. perturbation,
- ω = angular frequency of A.C. perturbation.

We define as the centre of the shock front the point where the particle concentration is the mean of the concentrations before and behind the shock, that is, where

$$n = \frac{1}{2}(n_0 + n_0').$$

* The basic equation (1) used in this paper has been derived in a paper on "The non-Maxwellian distribution in a shock front and the anomaly of the chromospheric temperature" (Sen 1953).

† The suffix e will be used when the concentration refers specifically to the electron.

For simplicity, we shall regard the shock centre as the representative point in the shock front and investigate the conditions at this point. Then the normalized electron velocity distribution function $f(U')$ at the shock centre, referred to axes moving with the stream velocity \bar{u} , can be shown to be (Sen 1953)

$$f(U') = \frac{1}{2\sqrt{\pi}} \frac{n_0}{n} \{e^{-(U'-f)^2} + de^{-b(U'+c)^2}\}, \quad \dots\dots\dots (1)$$

where b , c , d , and f are constants given by

$$\left. \begin{aligned} b &= \frac{(a-1)^2 M^2}{(M^2+a-2)(aM^2-1)}, \\ c &= \sqrt{\frac{a(a-2)}{2}} \frac{(M^2-1)(M^2+a-2)}{M(a-1)\{a(M^2+1)-2\}}, \\ d &= \frac{(a-1)^2 M^3}{(M^2+a-2)^{3/2}(aM^2-1)^{1/2}}, \\ f &= \sqrt{\frac{a(a-2)}{2}} \frac{M(M^2-1)}{a(M^2+1)-2}. \end{aligned} \right\} \dots\dots\dots (2)$$

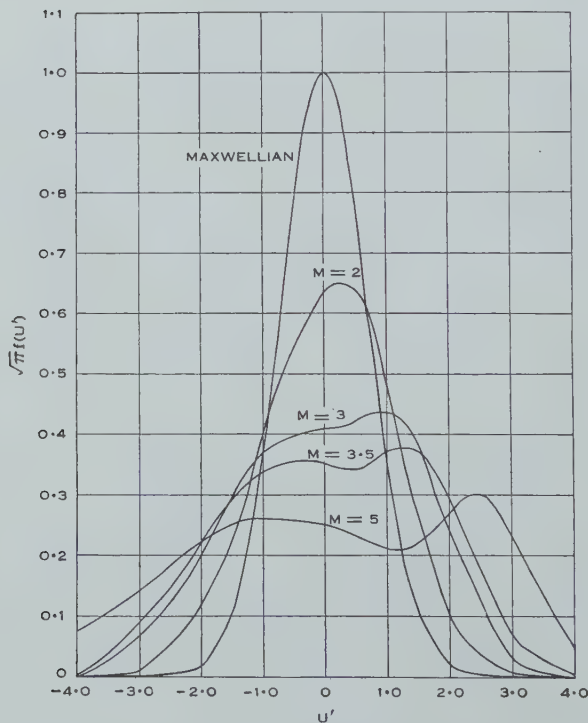


Fig. 1.—Velocity distribution at centre of shock front.

We take $\gamma=5/3$ for the electronic gas. Figure 1 gives the plot of the function $\sqrt{\pi}f(U')$ against U' for different values of the Mach number M . For

comparison, we have also drawn in the same figure the corresponding Maxwellian distribution.* It is apparent that, with increasing strength of shock, the velocity distribution curve becomes progressively non-Maxwellian. For $M \approx 3$, the curve develops a secondary hump that accentuates with M .

The secondary humps in our electron velocity distribution curves in Figure 1 give us reason to expect space charge wave amplification in the shock front. We shall see if this is so from the dispersion equation† corresponding to the distribution function (1). Assuming that all A.C. quantities vary as $Re \exp j(kx - \omega t)$, Vlasov (1945) derived, in the linear approximation, the following dispersion equation :

$$\frac{4\pi n_e e^2}{m} \int \frac{f(V_0) dV_0}{(\omega - kV_0)^2} = 1, \quad \dots \dots \dots (3)$$

where $f(V_0)$ is the normalized electron velocity distribution function.

Applying (3) to (1) and assuming a series expansion for small k (i.e. long waves), we derive the following dispersion equation :

$$\left(\frac{3c^2 d}{2\sqrt{b}} + \frac{3d}{4b^{3/2}} + \frac{3}{2}f^2 + \frac{3}{4} \right) k'^2 + \frac{d}{2\sqrt{b}} + \frac{1}{2} - \alpha^2 = 0, \quad \dots \dots (4)$$

where

$$k' = k/\omega \text{ and } \alpha = \omega/\omega_0, \quad \dots \dots \dots (5)$$

ω_0 being the plasma frequency in the medium before the shock, that is,

$$\omega_0^2 = \frac{4\pi n_0 e^2}{m}. \quad \dots \dots \dots (6)$$

Table 1 gives the upper limit of α as a function of M , for complex roots‡ of k' of the dispersion equation (4). Figure 2 gives the corresponding graph.

TABLE 1
UPPER LIMIT OF α AS A FUNCTION OF M

M	1.0	1.2	2.0	2.5	3.0	3.5	4.0	5.0	7.0	10.0
α	1.00	1.07	1.28	1.36	1.41	1.45	1.48	1.51	1.54	1.56

* $e^{-U^2} = \lim_{M \rightarrow 1} \sqrt{\pi} f(U).$

† We consider only the motions of the electrons and neglect the motions of the ions, on account of the relatively larger mass of the latter.

‡ It is true that the roots of k' in the equation as it stands are purely imaginary. The reason lies in the nature of the approximation and the neglect of collisions. Inclusion of higher powers of k' in (4) or of collision effects will introduce the real part of the propagation constant k .

Presumably, only values of $\alpha > 1$ will be significant for space charge wave amplification, as frequencies below the plasma frequency will not have much chance of escape from the overdense atmosphere. We see from Figure 2 that the frequency bandwidth of amplification becomes narrower with decreasing strength of shock.

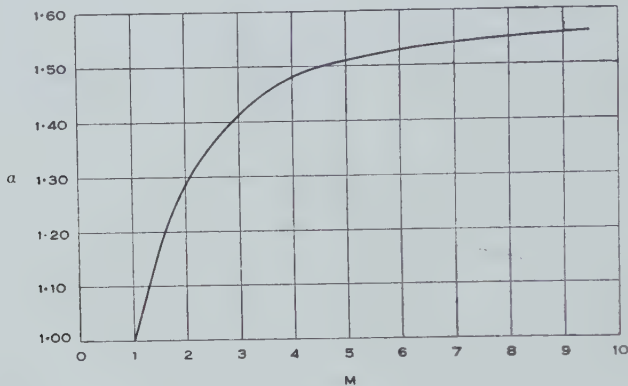


Fig. 2.—Frequency bandwidth of amplification, α , as a function of the shock strength M .

III. DISCUSSION

Frequency bandwidths of a few per cent. of the base frequency have been observed in the fine structure of solar radio noise bursts (Blum and Denisse 1950; Reber, unpublished data; Wild 1951). It is true that the Mott-Smith analysis does not apply to weak shocks.* Nevertheless, in the absence of any compelling reasons to suspect sharp discontinuities, the progressive decrease of the bandwidth of amplification with the shock strength would presumably carry over into the region of weak shocks, and is, moreover, in agreement with the analysis of Denisse and Rocard (1951).

We may suppose that the shock waves originate as weak shocks from the convective cells in the subphotospheric layers that are responsible for the photospheric granules (Schwarzschild 1948). The author (Sen 1953, Appendix) has shown that even in the low chromosphere the shock waves will be frequent enough to preserve the non-Maxwellian distribution as a quasi-steady state against the disruptive effect of collisions.

It is beyond the scope of the present paper to enter into the controversial and difficult question of whether the conditions are as favourable in the solar atmosphere as in discharge tubes for conversion of space charge wave energy into electromagnetic radiation.

The author believes, however, that the discontinuity at the shock front may favour such conversion. In that case, the shock fronts traversing the solar atmosphere would give rise to storm bursts of narrow bandwidth that form the elementary fine-structure components of solar radio noise bursts.

* It gives the wrong slope to the shock thickness as a function of the shock strength. See Wang Chang (1948) and Mott-Smith (1951).

IV. ACKNOWLEDGMENT

The author wishes to acknowledge his indebtedness to Miss Loris B. Perry for computing Table 1 and drawing the figures.

V. REFERENCES

- BLUM, E. J., and DENISSE, J. F. (1950).—*C.R. Acad. Sci. Paris* **231**: 1214.
BOHM, D., and GROSS, E. P. (1949).—*Phys. Rev.* **75**: 1851.
CHAPMAN, S., and COWLING, T. G. (1939).—“The Mathematical Theory of Non-Uniform Gases.” (Cambridge Univ. Press.)
DENISSE, J. F., and ROCARD, Y. (1951).—*J. Phys. Radium* **12**: 893.
LANDAU, L. (1946).—*J. Phys. Moscow* **10**: 25.
MOTT-SMITH, H. M. (1951).—*Phys. Rev.* **82**: 885.
SCHWARZSCHILD, M. (1948).—*Astrophys. J.* **107**: 1.
SEN, HARI K. (1953).—*Phys. Rev.* **92**: 861.
VLASOV, A. (1945).—*J. Phys. Moscow* **9**: 25, 130.
WANG CHANG, C. S. (1948).—On the theory of the thickness of weak shock waves. Univ. Michigan, Dept. Engng. Rep. UMH-3-F(APL/JHU CM-503).
WILD, J. P. (1951).—*Aust. J. Sci. Res.* **A4**: 36.

ON THE CRITERION FOR DIELECTRIC BREAKDOWN IN IONIC CRYSTALS

By J. J. O'DWYER*

[Manuscript received September 25, 1953]

Summary

Previous theories on the criterion for electronic breakdown in ionic crystals are reviewed, in particular those of Fröhlich and Heller. It is shown that the ideas proposed by Heller are based on insufficient examination of the physical picture of breakdown, and the validity of Fröhlich's criterion is vindicated so long as electron-lattice collisions determine the distribution function. However, for the relatively high densities of conduction electrons existing just before breakdown the effect of electron-electron collisions may well be more important than the effect of electron-lattice collisions in determining the distribution function. Accordingly a critical field strength is calculated for a Maxwellian distribution of conduction electrons by a method which is a modification of Fröhlich's work. Comparison with the somewhat inadequate experimental results existing at present suggests that the breakdown field should be identified with the critical field originally proposed by Fröhlich.

I. INTRODUCTION

The theory of the dielectric breakdown of ionic crystals has been dealt with in detail by Fröhlich (1937, 1939), who developed a criterion which could be applied for the calculation of the breakdown field strength at sufficiently low temperatures.

Recently Callen (1949), Heller (1951), and Franz (1952) have advanced criteria for dielectric breakdown in certain cases. The work of Callen is based on the so-called low energy criterion of von Hippel (1935) as opposed to Fröhlich's high energy criterion, and will not be discussed further since it has been dealt with in the literature and is, moreover, not required for the line of thought of the present work. Heller uses an idealized model of a non-polar crystal to calculate in detail the distribution function for the conduction electrons at high field strengths. The results of this calculation are then incorporated in his breakdown criterion to estimate the breakdown field strength. The actual numerical work is done for diamond, but Heller states that his criterion is also applicable to ionic crystals and compares the results of his criterion with those of Fröhlich. Franz likewise endeavours to calculate this distribution function but uses a method which is incorrect, since he assumes that the effect of the field is simply to shift the electron distribution function bodily in momentum space by an amount determined by some relaxation time constant. This has been shown to be in error for high fields in insulators by Fröhlich (1947b).

* Division of Electrotechnology, C.S.I.R.O., University Grounds, Sydney.

In the present work breakdown criteria are re-examined and it is found that the original Fröhlich criterion is correct for the model to which it was applied. However, it is possible that the model is not in accord with reality on some points; in particular, evidence will be advanced to show that electron-electron collisions may be much more frequent than electron-lattice collisions just before breakdown. A modification of Fröhlich's calculations (1937, 1947a) will be applied to determine a critical field strength for the instability of a Maxwellian distribution of conduction electrons.* Experimental evidence, though not convincing, will be found to be slightly in favour of the original Fröhlich theory.

II. BREAKDOWN CRITERIA

(a) *The Fröhlich Criterion*

Fröhlich (1939, 1947a) has distinguished two extreme cases for the establishment of equilibrium in electron distributions at high field strengths. In the first instance he has considered that collisions between conduction electrons and lattice vibrations maintain equilibrium together with the applied field, while collisions between conduction electrons themselves are of negligible consequence. This corresponds to the so called "low temperature" breakdown theory. Alternatively, it may be considered that collisions amongst conduction electrons themselves and between conduction electrons and electrons in shallow traps are much more frequent than collisions between conduction electrons and lattice vibrations. This assumption leads to the "high temperature" breakdown theory.†

In the case of the low temperature theory Fröhlich (1937) has calculated the mean rate of energy gain from the field as

$$A(F, E) = \frac{e^2 F^2}{m} \tau(E), \quad \dots \dots \dots (1)$$

where

$$\left. \begin{aligned} \frac{1}{\tau(E)} &= \frac{1}{\tau_0(E)} \left\{ 1 + \frac{2}{\exp(h\nu/kT_0) - 1} \right\}, \\ \frac{1}{\tau_0(E)} &= \frac{2^{2/3} \pi}{16\sqrt{2}} \cdot \frac{e^4 h}{m^{\frac{1}{2}} M a^5 \nu} E^{-3/2}, \end{aligned} \right\} \quad \dots \dots \dots (2)$$

and the mean rate of energy loss to the lattice vibrations by an electron of energy E as

$$B(E) = \frac{2^{\frac{1}{2}} \pi e^4 m^{\frac{1}{2}}}{M a^3} \log \gamma E^{-\frac{1}{2}}, \quad \dots \dots \dots (3)$$

* Since this paper was written the author has been informed by Professor Fröhlich that he has carried out the calculation of the critical field strength for a Maxwellian distribution of conduction electrons. It appears that his work is quantitatively more exact than that presented here, so that he may be able to draw more definite conclusions in his comparison with experimental work than those drawn on the basis of the present calculations.

† This distinction concerns the mechanism by which the electron distribution function is stabilized. It will be found necessary to make a further distinction below based on the mechanisms of energy transfer from the electrons to the lattice. This will lead to yet another set of physical conditions for which an instability criterion can be given for the electron distribution function.

in which e and m are the electronic charge and mass respectively, F is the electric field strength, ν the reststrahlen frequency, M the reduced ionic mass, a the lattice constant, T_0 the lattice temperature, and $\log \gamma$ is a slowly varying function of E such that, for $E=I$ (ionization energy), $(\log \gamma)^{\frac{1}{2}}=2.6$ for most alkali halides. These formulae are valid for energies greater than a certain energy ($E_0 \sim 1$ eV) and less than the ionization energy ($I \sim 5$ eV).

Fröhlich's low temperature breakdown criterion is then

$$A(F, I) = B(I), \quad \dots \dots \dots (4)$$

which is used as a determining equation for the breakdown field strength F^* . The physical idea of this criterion can be explained with the aid of Figure 1.

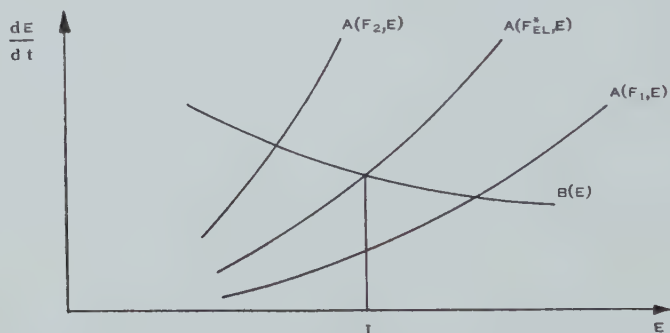


Fig. 1.—Average rate of energy gain from the field and energy loss to the lattice vibrations per electron.

For a field $F_1 < F^*$ the only electron which can, on the average, make a net energy gain from the field will be one whose energy is greater than the ionization energy; while for a field $F_2 > F^*$ an electron of less than ionization energy will make a net average gain from the field. It is then argued that F^* is the breakdown strength, since it is at this field strength that the field is, on the average, first able to accelerate an electron of slightly lower energy up to the ionization energy.

(b) *The Heller Criterion*

Concentrating essentially on a low temperature model (in the sense explained above), Heller (1951) has modified Fröhlich's idea and produced a new breakdown criterion. Remarking that Fröhlich has only considered the average behaviour of one electron, Heller rightly points out that high field strengths will change the electron distribution function from a Maxwellian one (as already shown by Fröhlich (1947b)) and then advances a breakdown criterion based on the idea that the field alters the ionization rate by altering the distribution function. He alleges that Fröhlich has ignored this by simply considering the change which the field produces in the average ionization process caused by one electron.

The essence of his work can most easily be seen by considering the kinetic equation derived by Fröhlich (1947b) :

$$\left. \begin{aligned} \frac{\partial f_0}{\partial t} &= \frac{B(x)}{kT} \frac{\partial}{\partial x} \left[\{1 + C^2(x)\} \frac{\partial f_0}{\partial x} + f_0 \right], \\ f_1 &= -\sqrt{\frac{2}{mkT_0}} eF\tau(x) \frac{\partial f_0}{\partial x}, \end{aligned} \right\} \dots\dots\dots (5)$$

in which we have defined a dimensionless variable $x = E/kT$, and expanded the electron distribution function in a series of Legendre polynomials

$$f(x, \theta, t) = f_0(x, t) + f_1(x, t) \cos \theta,$$

assuming that only these two terms are required. (θ is the angle between the applied field and the momentum direction and t is the time). The quantities $A(x)$, $B(x)$, and $\tau(x)$ are given by (1), (3), and (2) respectively, and

$$C^2(x) = \frac{2}{3} \cdot \frac{A(x)x}{B(x)}.$$

The equations (5) are valid only in the same energy range as equations (1), (2), and (3). The first equation of (5) can be written

$$\frac{\partial \rho}{\partial t} = \frac{\partial S}{\partial x}, \dots\dots\dots (6)$$

where ρ is the density function

$$\rho(x, t) = 4\pi\sqrt{x} f_0(x, t),$$

and S is a current density,

$$S(x, t) = \frac{4\pi\sqrt{x} B(x)}{kT} \cdot \left[\{1 + C^2(x)\} \frac{\partial f_0}{\partial x} + f_0 \right]. \dots\dots\dots (7)$$

For stationary conditions we have then from (6) that

$$S(x) = \text{const.} \dots\dots\dots (8)$$

Fröhlich has imposed the boundary condition that there can be no diffusion of electrons into or from the negative energy region and hence (8) becomes

$$S(x) = 0. \dots\dots\dots (8a)$$

This is simply a mathematical statement of the fact that, under the assumptions made, there are no electron removal or introduction processes.

However, it is clear that such processes will be important particularly at high field strengths. The ionization rate will depend on the field strength through the field dependence of both the distribution function and the total number of conduction electrons. If we denote the ionization rate by $S(F)$ then Heller replaces (8a) by

$$S(x) = S(F), \dots\dots\dots (8b)$$

alleging that, in the steady state, the ionization rate must be equal to the recombination and trapping rate. Since recombination and trapping occur mainly from amongst the low energy electrons, equation (7) will still determine the electron distribution function for energies between E_0 and I (the range of validity for equations (2) and (3)), in spite of the fact that electron-electron collisions are now being considered to be of importance in some energy ranges. Equation (7) then becomes

$$\{1 + C^2(x)\} \frac{\partial f_0}{\partial x} + f_0 = \frac{kT}{4\pi\sqrt{x}B(x)} S(F), \quad \dots\dots\dots (9)$$

and we note from (3) that $\sqrt{x}B(x)$ is independent of x . The solution to (9) is easily obtained as

$$f_0 = \frac{kT}{4\pi\sqrt{x}B(x)} S(F) + N \exp\left\{-\int_0^x \frac{dy}{1+C^2(y)}\right\}, \quad \dots\dots\dots (10)$$

in which N and S are to be determined by the normalization condition and the details of the ionization process.

For low energies or very low field strengths the solution (10) becomes

$$f_0 = \frac{kT}{4\pi\sqrt{x}B(x)} S(F) + N_1 \exp(-x), \quad \dots\dots\dots (10a)$$

while for higher field strengths and higher energies

$$f_0 = \frac{kT}{4\pi\sqrt{x}B(x)} S(F) + N_2 \exp(\Delta F^{-2}x^{-2}), \quad \dots\dots\dots (10b)$$

where Δ is a constant containing various factors appearing in $C^2(x)$. This solution (10b) is identical with the solution given by Heller for the same energy region, apart from different lattice interaction parameters which lead to a different value of Δ . Heller has performed very detailed calculations to determine the electron distribution function in the ionization region, for which purpose he has made arbitrary assumptions concerning the cross section for ionizing collisions. He then states that

$$\begin{aligned} S(F) &\equiv \alpha(F)N_C \\ &= AN_C\{N_T(F) + N_H\}, \quad \dots\dots\dots (11) \end{aligned}$$

where N_C is the number of electrons in the conduction band, $N_T(F)$ the number of empty trapping centres at field F , N_H the number of holes in the valence band, and A must apparently be understood as some sort of combined recombination and trapping coefficient. For an insulator obviously

$$N_C = N_H + N_T(F) - N_T(0), \quad \dots\dots\dots (12)$$

since conduction electrons can come only from the valence band or from traps. Combining (11) and (12) we have

$$\alpha(F) = A\{N_C + N_T(0)\}. \quad \dots\dots\dots (13)$$

Heller then gives as his breakdown criterion

$$\alpha(F) \approx A \cdot N_T(0), \quad \dots\dots\dots (14)$$

on the ground that N_c begins to increase rapidly (proportionally to $\alpha(F)/A$) when $\alpha(F)$ exceeds $A \cdot N_T(0)$. This means that breakdown is to be considered as occurring when the ionization rate is so high that the trapping rate cannot keep pace with it, the equilibrium number of conduction electrons then being such a rapidly increasing function of field strength that no dielectric would carry the current involved.

(c) *A Re-examination of the Breakdown Criteria*

Although, at first sight, Heller's ideas would appear to be a logical extension of Fröhlich's work, nevertheless it seems that they are based on insufficient examination of the physical aspects of the problem. The main misconception in his treatment is that he assumes that the ionization rate will be balanced by the trapping and recombination rate when the conduction electrons have attained their equilibrium distribution in a given external field. This cannot be so; for, if trapping and recombination were together responsible for balancing ionization from the valence band, then a net transfer of electrons from the valence band to traps would be continuously required. There could not then be any equilibrium no matter how weak the field. It is plain that, in the state of equilibrium, ionization from the valence band must be balanced by recombination to the valence band and ionization from traps must be balanced by trapping processes. (Ionization in both these cases may be due to thermal ionization, ionization by inelastic collision with conduction electrons, and field emission. In the case of traps all three processes will probably contribute, but ionization from the valence band will be mainly due to inelastic collisions.) It is now clear that it is incorrect to replace (8a) by (8b), since a given ionization rate $S(F)$ does not imply a current of electrons in energy space, as each high energy electron lost by an ionizing collision will be replaced by one from a recombining collision. Equations (9) and (10) will then be correct with $S(F)=0$ (as originally given by Fröhlich), and the breakdown field would logically be conceived as that field for which the recombination rate is unable to balance the ionization rate. However, this is exactly what Fröhlich has calculated in his original paper. It is easily seen that for fields $F \gg F^*$ the high energy electron from a recombining collision is not, on the average, able to lose its energy to the lattice, and in fact undergoes a further ionizing collision thus nullifying the effect of recombination in balancing ionization. Thus the Fröhlich criterion gives the field strength for which recombining collisions are unable to effect a net removal of electrons from the conduction levels.

(d) *A New Approach to the Breakdown Criterion*

It would seem that the breakdown field strength of ionic crystals in the low temperature region should be identified with that critical field strength given by Fröhlich, since at this field strength ionization would increase the number of conduction electrons indefinitely. However, it seems to have been overlooked that for large electron densities electron-electron collisions will be the dominating feature in determining the distribution, and not electron-lattice collisions as assumed by both Fröhlich and Heller in their derivation of the

distribution function.* Moreover, since the time taken for conduction electrons to establish a temperature under the influence of mutual collisions is many orders of magnitude less than the time required for breakdown to build up, the electron distribution can be expected to change back to Maxwellian under these conditions. This process decreases the number of electrons of high energy (of order of the ionization energy) to such an extent that, while the ionization rate is not actually zero, it is of such infinitesimal proportions that it could scarcely be conceived as being responsible for breakdown (cf. (10a) and (10b)).

The critical field strength is then that field strength for which a Maxwellian distribution of conduction electrons cannot attain any finite temperature for which equilibrium will exist between the energy gain from the field and the energy loss to the lattice by the conduction electrons. This picture, which is at first sight similar to the model on which Fröhlich (1947a) calculates his high temperature breakdown, differs from it in very important respects :

(i) In the low temperature case the number of conduction electrons will be independent of their temperature T for any given lattice temperature T_0 . The number will in fact be that for which the electron-electron collision relaxation time has just superseded the electron-lattice collision relaxation time as being the chief factor determining the distribution function. In the high temperature case, on the other hand, the number of conduction electrons is a very sensitive function of T , due to excitation from the large numbers of shallow traps which exist at the higher temperatures.

(ii) In the low temperature case the transfer of energy from the electrons to the lattice is almost entirely due to collisions between conduction electrons and lattice vibrations, while in the high temperature case practically all the energy lost to the lattice is by transitions of electrons in shallow traps.

III. THE CRITICAL FIELD STRENGTH

The low temperature critical field strength will thus be determined from

$$A(F, T, T_0) = B(T, T_0), \quad \dots \dots \dots (15)$$

as the lowest value of F for which (15) yields no finite solution for T . (That this is, in fact, possible is evident from the analysis below.) Since the number of conduction electrons is not a function of T we have

$$A(F, T, T_0) = \frac{e^2 F^2}{m} \tau(T, T_0), \quad \dots \dots \dots (16)$$

* Heller has estimated the relative importance of electron-electron collisions by asserting that 10^{12} conduction electrons per c.c. give evidence of breakdown. This estimate seems unreasonably low, particularly since it is implicit in his breakdown criterion that the density of conduction electrons just before breakdown is of the same order of magnitude as the number of electron acceptor impurity atoms, which figure he estimates at 10^{17} per c.c. This latter figure is more in agreement with the order of magnitude obtained experimentally on mica by Kawamura, Onuki, and Okura (1952) who found that avalanches contained about 10^8 electrons just before breakdown. Since their specimens were 5×10^{-4} cm thick, any reasonable estimate of the "avalanche cross section" will give a figure more of order 10^{17} than 10^{12} for the electron density. It is for just such densities that electron-electron collisions would be expected to achieve the same degree of importance as electron-lattice collisions.

in which $\tau(T, T_0)$ is the average relaxation time given by

$$\tau(T, T_0) = \frac{\int \tau(E, T, T_0) \sqrt{E} \exp(-E/kT) dE}{\int \sqrt{E} \exp(-E/kT) dE} \quad (17)$$

The quantity $\tau(E, T, T_0)$ represents the average relaxation time of an electron of energy E in a Maxwellian distribution of temperature T with lattice temperature T_0 . It can be derived easily from Fröhlich's (1937) calculations as

$$\begin{aligned} \frac{\exp(-E/kT)}{\tau(E, T, T_0)} &= \frac{2a^3 N}{(2\pi)^3} \int_0^{2\pi} d\varphi \int_{-\pi}^{\pi} \sin \theta d\theta \\ &\times \int_0^{w_0} \frac{w^2}{2k^2} \left[\Phi_w^e \exp\{-(E+h\nu)/kT\} + \Phi_w^a \exp(-E/kT) \right] w^2 dw, \\ &\dots\dots\dots (18) \end{aligned}$$

in which a is the lattice constant, w is the wave number of the lattice wave, k is the wave number of the electron, and Φ_w^a , Φ_w^e are the probabilities per second that an electron makes a transition into or from the state considered by emission or absorption respectively. The integrations (representing the sum over all processes) have been performed by Fröhlich and give

$$\frac{1}{\tau(E, T, T_0)} = \frac{1}{\tau_0(E)} \frac{\exp(h\nu/kT_0 - h\nu/kT) + 1}{\exp(h\nu/kT_0) - 1}, \quad (19)$$

in which

$$\frac{1}{\tau_0(E)} = \frac{2\pi e^4 m^{\frac{1}{2}}}{2^{\frac{1}{2}} \hbar M a^3 \nu E^{\frac{1}{2}}} \quad (20)$$

replaces the expression for $\tau_0(E)$ given in (2) since we are now dealing with energies less than the critical energy stipulated. Substituting (19) and (20) in (17) and introducing the result into (16) gives

$$A(F, T, T_0) = \left(\frac{2}{m}\right)^{3/2} \frac{M a^3 h \nu}{2 e^2 (\pi)^{3/2}} F^2 \sqrt{kT} \frac{\exp(h\nu/kT_0) - 1}{\exp(h\nu/kT_0 - h\nu/kT) + 1} \quad (21)$$

Similarly the average loss to the lattice can be found from*

$$B(T, T_0) = \frac{\int B(E, T, T_0) \sqrt{E} \exp(-E/kT) dE}{\int \sqrt{E} \exp(-E/kT) dE} \quad (22)$$

* The averaging in (17) and (22) is not strictly correct. Fröhlich (personal communication) has pointed out that (17) should read

$$\tau(T, T_0) = \frac{-\int \tau(E, T, T_0) E^{3/2} \{\partial \exp(-E/kT) / \partial E\} dE}{\int E^{1/2} \exp(-E/kT) dE}$$

This will alter the result for $\tau(T, T_0)$ by the numerical factor 2, and, as no conclusions are drawn from the absolute magnitude of the critical field, the error will not affect the conclusions of this work.

The integrations of (17) and (22) have been taken from 0 to ∞ . Those in (17) should be from 0 to I and the form of the integrand should change for the critical energy ~ 1 eV. However, contributions from such relatively high energies will be small due to the form of the distribution function. The lower limit of the integrations in (22) should be $h\nu$ as electrons with energy less than $h\nu$ cannot lose energy to the lattice. This will probably be a more serious source of error.

Using Fröhlich's (1937) calculations we find similarly

$$B(E, T, T_0) = \frac{h\nu}{\tau_0(E)} \log \Gamma \frac{\exp(h\nu/kT_0 - h\nu/kT) - 1}{\exp(h\nu/kT_0) - 1}. \quad (23)$$

Substituting (23) in (22) and making use of (20) gives then

$$B(T, T_0) = \left(\frac{2m}{\pi}\right)^{\frac{1}{2}} \frac{2\pi e^4}{Ma^3} \frac{\log \Gamma}{(kT)^{\frac{1}{2}}} \frac{\exp(h\nu/kT_0 - h\nu/kT) - 1}{\exp(h\nu/kT_0) - 1}, \quad (24)$$

in which $\log \Gamma$ has been assumed to be independent of E (actually it is a slowly varying function of E of order of magnitude unity). The quantities $A(F, T, T_0)$ and $B(T, T_0)$ are plotted diagrammatically in Figure 2 from which it appears

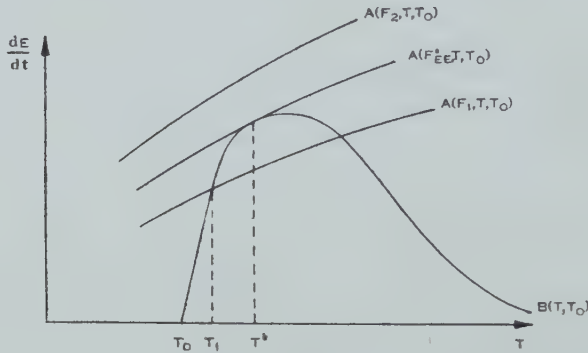


Fig. 2.—Average rate of energy gain from the field and energy loss to the lattice vibrations per electron of a Maxwellian distribution.

that for a field $F_1 < F^*$ an equilibrium temperature T_1 will be attained, while for a field $F_2 > F^*$ no equilibrium temperature for the conduction electrons will be possible. Hence F^* is the critical field strength for which the Maxwellian distribution of conduction electrons becomes unstable, the corresponding electronic temperature being T^* . An approximate analytic expression for F^* can be obtained subject to certain simplifying assumptions. Since the curves $A(F, T, T_0)$ are reasonably flat in the vicinity of $T = T^*$, we may take T^* to be that value which maximizes $B(T, T_0)$ with respect to T . The condition for a maximum is

$$(1 - h\nu/kT) \exp(h\nu/kT_0 - h\nu/kT) = 1, \quad (25)$$

and, if we assume $h\nu \leq kT_0$ (a condition which is satisfied for many of the alkali halides between about -100°C and room temperature), then $(h\nu/kT_0 - h\nu/kT) < 1$, and on expanding the exponential we find

$$T^* \simeq 2T_0. \quad (26)$$

Substitution of this value of T in the instability criterion (15) gives with the help of (21) and (24)

$$F_{EE}^* \simeq \frac{m\pi e^3}{Ma^3 h\nu} (\log \Gamma)^{\frac{1}{2}} \frac{[h\nu/kT_0 \{\exp(h\nu/kT_0) + 1\}]^{\frac{1}{2}}}{\exp(h\nu/kT_0) - 1}, \quad (27)$$

as an approximate expression for the critical field strength which causes instability when electron-electron collisions determine the distribution function of the conduction electrons. This will be compared with the value

$$F_{EL}^* = \frac{2^{1/3} m^{1/2} \pi \hbar e^3}{4 M a^4 I (\hbar \nu)^{1/2}} (\log \gamma)^{1/2} \left\{ 1 + \frac{2}{\exp (\hbar \nu / k T_0) - 1} \right\}^{1/2}, \quad \dots (28)$$

obtained originally by Fröhlich (1937) for the critical field strength which causes instability when electron-lattice collisions determine the distribution function of the conduction electrons.

IV. DISCUSSION

(a) Identification of the Breakdown Field

The question as to the identification of the breakdown field strength immediately arises, and it seems that breakdown will occur for that field strength which is the higher of F_{EE}^* or F_{EL}^* in any given instance.

For suppose $F_{EE}^* < F_{EL}^*$; then not until $F = F_{EL}^*$ would there be any mechanism by which the number of conduction electrons could be sufficiently increased to make electron-electron collisions predominate in determining the distribution function. However, when $F = F_{EL}^*$ the number of conduction electrons increases rapidly till the distribution becomes Maxwellian, and since $F_{EL}^* > F_{EE}^*$ breakdown should then occur.

If on the other hand $F_{EL}^* < F_{EE}^*$, it may reasonably be supposed that a field $F = F_{EL}^*$ simply changes the conduction electron distribution to a Maxwellian one whose temperature is about twice the lattice temperature (cf. equation (26)). For alkali halides with $I \gtrsim 5$ eV there would then be an almost negligible ionization rate. It is then reasonable to suppose that breakdown would not occur until $F = F_{EE}^*$.

(b) The Magnitude of the Breakdown Field

Quantitative calculations show that F_{EE}^* and F_{EL}^* are of the same order of magnitude for most of the alkali halides, and as given by (27) and (28) respectively are of the order of half the measured breakdown strengths given by von Hippel (1935). This is not as serious as would seem at first sight since many factors may be considered, some of which cause effects of considerable magnitude and all of which would tend to improve the agreement between theory and experiment.

Firstly, from the experimental point of view the results of von Hippel must be subject to some doubt. Recent careful work by Calderwood and Cooper (1953), in which the specimens were annealed so as to be as free as possible from mechanical strain, has shown that some of von Hippel's results are probably too high. Unfortunately, however, they have not tested a sufficiently full range of substances to make adequate comparison.

Secondly, from the theoretical point of view several comments arise. In the derivation of (27) the results of a perturbation calculation have been extended to regions of doubtful validity; also $\log \Gamma$ has been put as unity. Thus, granted

the model, less quantitative significance applies to the value of F_{EE}^* than to that of F_{EL}^* . Further, if the Born lattice theory of the dielectric constant of ionic crystals be accepted, then the reststrahlen frequency and several of the crystal constants appearing in the final results for F_{EE}^* and F_{EL}^* can be replaced by the experimentally measured dielectric constants, and the calculated breakdown strengths increased by factors of the order of 50 per cent. This has been done by Fröhlich (1939) for F_{EL}^* , but is not repeated for the present calculations, since no significance will be attached to the absolute magnitude of F_{EE}^* .

It is evident that in the present state of the theory an examination of the magnitudes of breakdown strengths will be of little use in determining which critical field strength is to be identified with the breakdown field strength.

(c) *Dependence of Breakdown Strength on Temperature*

Omitting universal constants and writing $z = \hbar\nu/kT_0$ we have for the critical field strengths

$$F_{EE}^* \propto \frac{1}{Ma^3\hbar\nu} \frac{[z\{\exp(z)+1\}]^{\frac{1}{2}}}{\exp(z)-1}, \quad \dots\dots\dots (29)$$

$$F_{EL}^* \propto \frac{1}{Ma^4I(\hbar\nu)^{\frac{1}{2}}} \left\{ 1 + \frac{2}{\exp(z)-1} \right\}^{\frac{1}{2}}. \quad \dots\dots\dots (30)$$

The temperature dependences are then contained entirely in the functions of z . The ratio of these temperature dependent terms is given by

$$\left\{ \frac{z}{\exp(z)-1} \right\}^{\frac{1}{2}} \simeq 1, \quad \dots\dots\dots (31)$$

the approximation being subject to the restriction imposed in the derivation of (27), namely, $z \leq 1$. It cannot be expected, then, that measurements of the temperature dependence will show which is the critical field strength to be identified with the breakdown field, since both criteria give approximately the same temperature dependence.

Calderwood and Cooper (1953), working on sodium chloride and potassium chloride in the temperature range -100 to 50°C , have found that the breakdown strength increases with increasing temperature more rapidly than predicted by the functions of (29) and (30). This is also understandable on the basis of Fröhlich's calculation of the temperature dependence of the relaxation time, since in this temperature range an increasing number of thermal defects (not taken into account in the calculations) would reduce the relaxation time at a faster rate than that predicted.

(d) *Dependence of Breakdown Strength on Physical Parameters*

The variation of breakdown strength at room temperature over a series of alkali halides is given in Table 1. The quantities listed are the experimental breakdown strength F_{EX}^* according to von Hippel (1935), the critical field F_{EL}^* according to Fröhlich (1937), and the critical field F_{EE}^* from equation (27).

Units have been adjusted so that all values given are relative to the corresponding value for sodium chloride. This is easily achieved for the theoretical results since from (27) and (28) we have immediately

$$\frac{F_{EE}^*}{F_{EL}^*} \propto \frac{aI}{(h\nu)^{\frac{1}{2}}} \cdot \left\{ \frac{z}{\exp(z)-1} \right\}^{\frac{1}{2}} \dots\dots\dots (32)$$

An examination of these results shows clearly that the breakdown strength varies from crystal to crystal in a manner which agrees much more closely with the critical field F_{EL}^* than with F_{EE}^* . Although it is possible that breakdown occurs at a field strength F_{EL}^* for some alkali halides and F_{EE}^* for others, and, although the accuracy of the experimental results quoted is dubious, it seems most likely on the basis of these experimental results that F_{EE}^* is less than F_{EL}^* for these substances, and that F_{EL}^* of Fröhlich's original work should be identified with the breakdown field strength.

TABLE I

RELATIVE VALUES OF BREAKDOWN FIELD STRENGTHS AND CRITICAL FIELD STRENGTHS
FOR A SERIES OF IONIC CRYSTALS AT ROOM TEMPERATURE

Substance	F_{EX}^*	F_{EL}^*	F_{EE}^*
Sodium chloride ..	1.00	1.00	1.00
Sodium bromide ..	0.66	0.99	1.15
Sodium iodide ..	0.53	0.87	1.01
Potassium chloride ..	0.66	0.62	0.86
Potassium bromide ..	0.47	0.52	0.77
Potassium iodide ..	0.40	0.48	0.70
Rubidium chloride ..	0.53	0.45	0.69
Rubidium bromide ..	0.40	0.39	0.68
Rubidium iodide ..	0.33	0.29	0.50

Heller has criticized Fröhlich's theory on the ground that it predicts that breakdown strength should vary inversely with internal ionization energy, and he suggests that two crystals with similar lattice parameters but different internal ionization energies be used to test the validity of his criticism. However, it is obvious that this criticism is quite invalid since the breakdown strength is a function of many other physical constants apart from the ionization energy which is contained implicitly in a knowledge of those constants together with the laws governing the formation of crystals. In other words, the best that any theory can do is to give a correct manner of variation from crystal to crystal, since in the nature of things it is not possible to vary one physical constant while keeping the others fixed.

V. ACKNOWLEDGMENTS

The author wishes to thank Professor H. Fröhlich, F.R.S., of the Department of Theoretical Physics, University of Liverpool, for helpful personal communications on this subject; also Dr. P. G. Klemens, Division of Physics, C.S.I.R.O., for stimulating discussions.

VI. REFERENCES

- CALDERWOOD, J., and COOPER, R. (1953).—*Proc. Phys. Soc. Lond.* B **66** : 74.
CALLEN, H. (1949).—*Phys. Rev.* **76** : 1394.
FRANZ, W. (1952).—*Z. Physik* **132** : 285.
FRÖHLICH, H. (1937).—*Proc. Roy. Soc. A* **160** : 230.
FRÖHLICH, H. (1939).—*Proc. Roy. Soc. A* **172** : 94.
FRÖHLICH, H. (1947a).—*Proc. Roy. Soc. A* **188** : 521.
FRÖHLICH, H. (1947b).—*Proc. Roy. Soc. A* **188** : 532.
HELLER, W. (1951).—*Phys. Rev.* **84** : 1130.
VON HIPPEL, A. (1935).—*Ergebn. exakt. Naturw.* **14** : 79.
KAWAMURA, H., ONUKI, M., and OKURA, H. (1952).—*J. Phys. Soc. Japan* **7** : 528.

THE INTEGRAL AND DIFFERENTIAL RANGE SPECTRA OF SEA-LEVEL MESONS

By A. J. DYER*

[*Manuscript received November 30, 1953*]

Summary

The differential and integral range spectra of the hard component of cosmic rays in water and lead have been determined up to a thickness of 1500 g cm^{-2} of water, and 2800 g cm^{-2} of lead. The differential results indicate that there is no anomaly with an intensity greater than 5 per cent. in the differential momentum spectrum in the region below 4 BeV/c. A comparison of the integral range and momentum spectra supports the energy loss data of Halpern and Hall in preference to those of Bethe and Bloch in the case of water. No such distinction of the theories is possible in the case of lead and, although the integral absorption curve in lead agrees with that obtained by Heyland and Duncanson (1953*b*), there is an unexplained discrepancy between experiment and theory. The intensity of knock-on showers as a function of thickness of water absorber has also been determined.

I. INTRODUCTION

In an earlier report (Dyer 1953) an account was given of a determination of the integral and differential range spectra of sea-level mesons in water. The differential data obtained did not possess sufficient accuracy to lead to a conclusion on the reality of an anomaly in the momentum spectrum near 3 BeV/c referred to by several authors (Blackett 1937 ; Glaser, Hamermesh, and Safonov 1950 ; Caro, Parry, and Rathgeber 1951).

As there was some evidence of an anomaly, the integral and differential spectra (in the neighbourhood of 4 BeV/c) were studied further with improved equipment. A greater reliability was obtained by replacing battery operated circuits with mains operated ones, adequately stabilized against voltage fluctuations.

Since the water-hole was available during the winter months only, corresponding measurements were made with a lead absorber during the summer. The results of both experiments are reported in this paper.

II. LEAD ABSORPTION EXPERIMENT

The general design of the counter telescope was largely determined by the fact that a quantity of lead in ingot form of the order of $1\frac{1}{2}$ tons was available on loan. In order to obtain the necessary absorber thickness only one arrangement of the lead was possible, that is, within the telescope with the cross-sectional dimensions of the absorber approximately 9 by 6 in. The differential absorber had a thickness of 20 cm.

* Physics Department, University of Melbourne.

The telescope arrangement is shown in Figure 1. Counters *A*, *B*, and *C* had an effective length of 18 cm, and counters in tray *D* had an effective length of 50 cm. The counters placed around trays *B* and *C* were in anti-coincidence

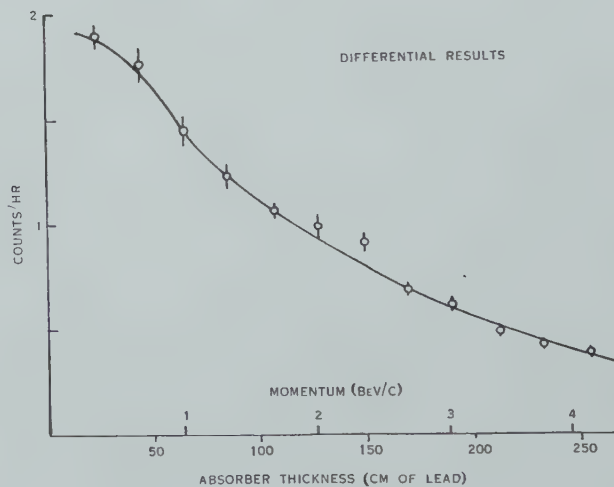
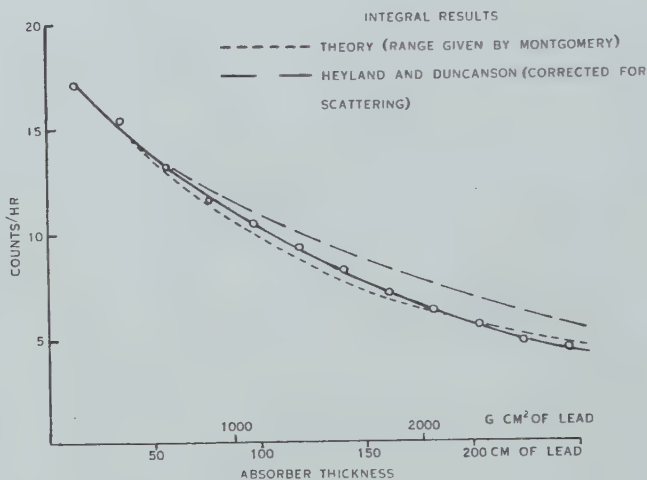
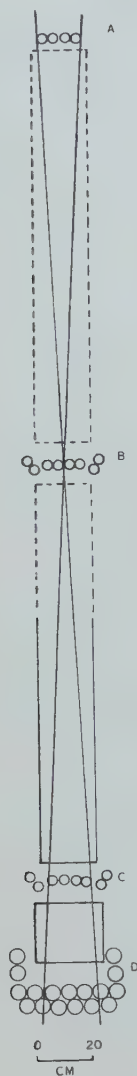


Fig. 1

Fig. 2

Fig. 1.—Counter arrangement for lead absorption experiment.

Fig. 2.—Integral and differential range spectra in lead.

with the threefold coincidence arrangement *ABC* and provided protection against side showers. Coincidences *ABC* and *ABC-D* were recorded for absorber thicknesses of from 14 cm up to 245 cm in steps of 21 cm. The coincidence circuits had a resolving time of 5 μ sec.

The anti-coincidence background rate was obtained by removing the differential absorber and a smoothed correction applied to the differential counting rates.

The final results are plotted in Figure 2 and presented in Table 1.

TABLE 1
INTEGRAL AND DIFFERENTIAL RANGE SPECTRA IN LEAD

Integral		Differential	
(counts/hr)	Range (g cm ⁻²)	(counts/hr)	Mean Range (g cm ⁻²)
17.2 ± 0.2	159	1.90 ± 0.06	272
15.4 ± 0.3	397	1.77 ± 0.08	511
13.2 ± 0.2	636	1.45 ± 0.07	749
11.6 ± 0.2	874	1.22 ± 0.06	987
10.5 ± 0.1	1112	1.06 ± 0.03	1226
9.4 ± 0.2	1350	0.99 ± 0.05	1463
8.2 ± 0.1	1589	0.92 ± 0.04	1702
7.1 ± 0.1	1828	0.68 ± 0.03	1941
6.3 ± 0.1	2065	0.61 ± 0.03	2179
5.7 ± 0.1	2305	0.48 ± 0.02	2419
4.9 ± 0.1	2545	0.42 ± 0.02	2657
4.6 ± 0.1	2780	0.38 ± 0.02	2895

III. WATER ABSORPTION EXPERIMENT

The new counter telescope used in conjunction with the underwater equipment described previously (Dyer 1953) is shown in Figure 3. The main features of the earlier arrangement were preserved, but larger counting rates were obtained by increasing the size of the counter trays.

Threefold coincidences *ABC* were recorded of particles traversing 10 cm of lead, and the number stopping in a further 10 cm of lead obtained by means of the double-layer anti-coincidence tray *D*. Two counters *E* were placed on either side of tray *B* in anti-coincidence with trays *ABC* to provide protection against side showers and showers produced in absorber *a*.

An additional tray of counters *FG* was placed immediately under absorber *a* with alternate counters connected in parallel. Coincidences *AFG* were selected and recorded and used also as an anti-coincidence "suppression" pulse on the coincidence arrangement *ABC*. Events *AFG* would be mainly knock-on showers and with this counter arrangement would be detected with an efficiency of about 50 per cent. The use of counters *EEG* in this manner provided adequate protection against unwanted events.

The counters *ABCEFG* had an effective length of approximately 18 cm and were 3 cm in diameter. The counters in tray *D* had an effective length of 40 cm and were 5 cm in diameter. All counters were of the Maze type and were constructed according to the technique of MacKnight and Chasson (1951).

The coincidence and amplifier circuits (resolving time 5 μsec) were located in the underwater tank and operated from stabilized B+ and heater voltages.

The main power supply and recording unit was situated in a hut adjoining the water-hole.

The integral and differential range spectra were determined down to a depth of 14 m of water in steps of approximately 1 m, and the anti-coincidence back-

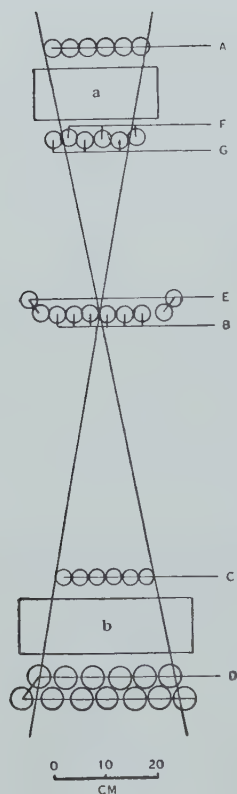


Fig. 3

Fig. 3.—Counter arrangement for water absorption experiment.

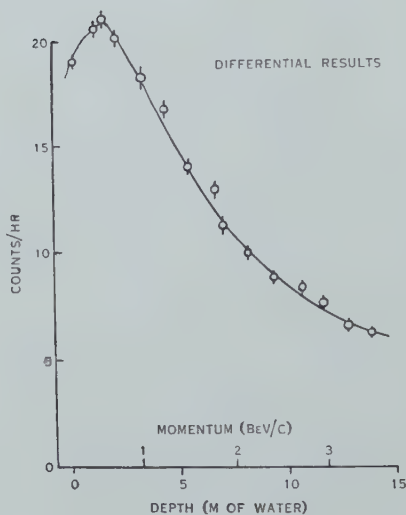
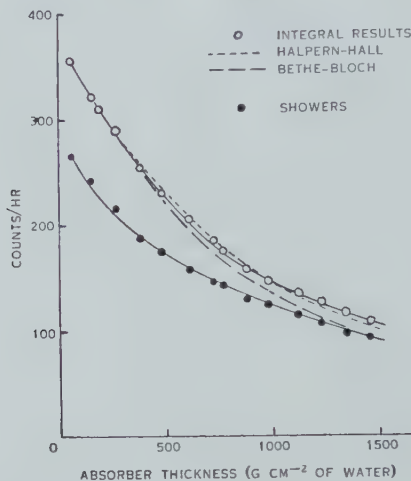


Fig. 4

Fig. 4.—Integral, differential, and shower curves in water.

ground measurement was made by removing the differential absorber. At the same time measurements of the shower intensity (AFG) were obtained over the same range of absorber thickness.

The results are presented in Figure 4 and in Table 2.

IV. DISCUSSION OF RESULTS

(a) *Integral Spectra*

The integral results are compared in Figure 5 with the curve given by Rossi (1948). In order to convert to equivalent range in g cm^{-2} of air use was made of the data given by Montgomery (1947) for the range of mesons in air and lead. The momentum-range relation for water was obtained by integrating the energy loss data given by Halpern and Hall (1948), and all results were normalized to Rossi's curve at 98 g cm^{-2} of air.

TABLE 2
RESULTS OF UNDERWATER INVESTIGATION

Depth (m of water)	Integral (counts/hr)	Differential (counts/hr)	Showers (counts/hr)
0	355 ± 1.3	19.0 ± 0.3	265 ± 1.3
1.03	321 ± 1.9	20.5 ± 0.5	241 ± 1.6
1.30	310 ± 1.7	21.0 ± 0.4	232 ± 1.0
2.12	290 ± 1.4	20.1 ± 0.4	216 ± 1.3
3.24	254 ± 1.6	18.3 ± 0.5	194 ± 1.3
4.29	230 ± 1.3	16.8 ± 0.4	177 ± 1.2
5.52	205.2 ± 1.1	14.0 ± 0.3	158.6 ± 1.2
6.70	185.1 ± 1.2	13.0 ± 0.4	147.0 ± 1.0
7.12	174.8 ± 1.2	11.2 ± 0.4	142.0 ± 1.1
8.19	159.0 ± 0.9	10.0 ± 0.2	129.0 ± 1.1
9.36	147.5 ± 1.0	8.9 ± 0.3	120.7 ± 1.0
10.72	136.2 ± 1.0	8.4 ± 0.3	112.3 ± 0.9
11.77	127.4 ± 0.7	7.6 ± 0.2	103.8 ± 0.7
12.90	116.2 ± 0.8	6.5 ± 0.3	97.3 ± 0.7
14.00	108.3 ± 0.6	6.1 ± 0.2	91.8 ± 0.6

The general agreement with the Rossi curve can be seen from Figure 5, and in particular it will be noted that the results for water agree somewhat better than those for lead. Rossi's curve is based on the work of Ehmert (1937) and Wilson (1938), and the present results lie within the experimental errors of their measurements.

(b) *Differential Spectra*

For the purpose of deciding the reality of an anomaly in the momentum spectrum, a differential measurement of the type described here possesses a greater degree of sensitivity than an integral one.

When both differential experiments are taken into account, it can be said with confidence that no anomalies greater than 5 per cent. exist in the momentum spectrum below 4 BeV/c .

This result agrees with that of Brini, Rimondo, and Filosofo (1952) who, from a similar experiment using an iron absorber, concluded that there are no irregularities greater than a few per cent. Similarly, Heyland and Duncanson (1953a, 1953b) decided by a detailed investigation using a lead absorber that there are no significant anomalies in the region 0–365 cm of lead.

By making use of the normalizing factors fitting the integral results to the corresponding Rossi curve, and the equivalent ranges in air and lead given by Montgomery (1947) it is possible to estimate the absolute intensity of the differential spectrum from the present results (Fig. 6). It is seen that the

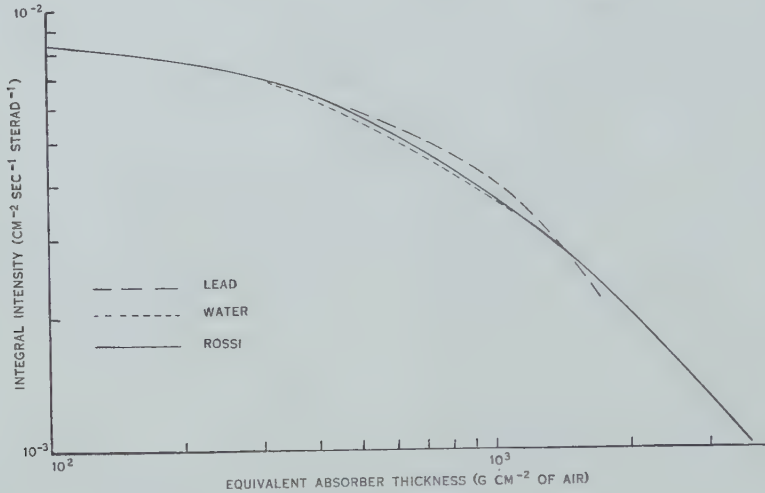


Fig. 5.—Integral range spectra in absolute units.

experimental curves are of similar shape to the Rossi differential curve but lie somewhat higher. This result is consistent with the suggestion offered by York (1952) that the Rossi curve is 20 per cent. too low.

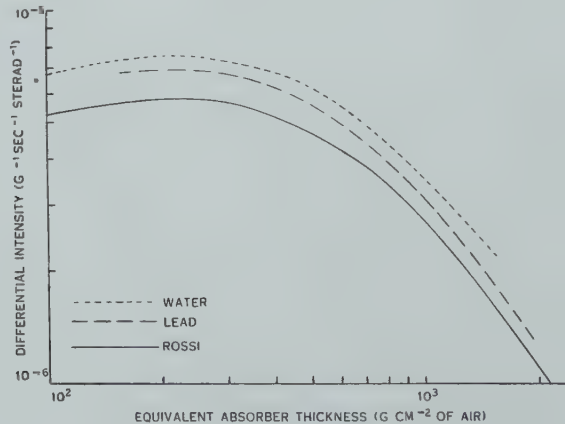


Fig. 6.—Differential range spectra in absolute units.

(c) Shower Measurements

The intensity of knock-on events as a function of thickness of water absorber has been presented in Figure 4. Two aspects of this shower measurement are of interest: (a) the ratio of the number of knock-on events to the meson intensity

and (b) the way in which this ratio varies with depth. Janossy (1948, p. 245) calculates the ratio of knock-on events at sea-level to be 0.10, and this agrees with the present results when it is noted that this type of counter arrangement has a detection efficiency of approximately 50 per cent. for two-particle showers.

The variation of this ratio with depth is shown in Figure 7 in which the value at zero depth has been normalized to the above value of 0.10. Two calculations of the variation of the intensity of knock-on events with depth underground have been made (Janossy 1948, p. 247; Hayakawa and Tomonaga 1949), but these authors are mainly concerned with very great depths and no profitable comparison with their results is possible.

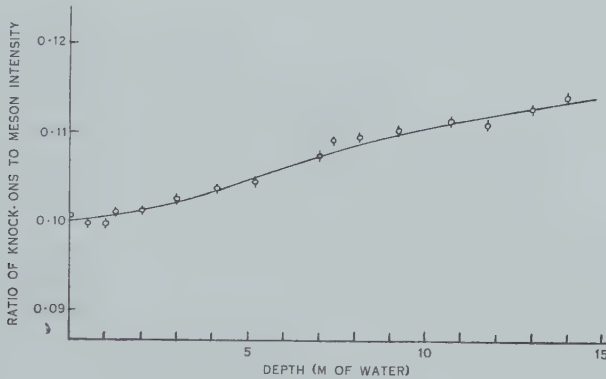


Fig. 7.—Ratio of knock-on showers to meson intensity as a function of depth.

V. RANGE-MOMENTUM RELATION FOR HIGH ENERGY MESONS

It is evident from the work of George (1952) that the only significant process causing energy loss for mesons with energy less than 4 BeV is that of ionization. Thus by comparing the intensity-depth curves with curves based on the integral momentum spectrum and theoretical energy loss data it should be possible to make some test of the different theories.

The chief point of interest lies in the difference between the familiar Bethe-Bloch energy loss curve and that given by Halpern and Hall (1948). The data for water provide a good test in this respect because of the large polarization effect predicted for water by Halpern and Hall. This is not so in the case of lead below 4 BeV/c and the difference between the two theories is insignificant for our purpose.

Thus in Figures 2 and 4 theoretical absorption curves are presented making use of the data given by Halpern and Hall (1948) for water, Montgomery (1947) for lead, and the integral momentum spectrum as determined by the Melbourne cosmic ray spectrometer. To permit normalization at the minimum absorber thickness the following information was used: a range of 14 cm of lead corresponds to 0.29 BeV/c; 10 cm of lead corresponds to 0.24 BeV/c and to 58 g cm⁻² of water.

Considering first the results for water, it is seen from Figure 2 that the experimental results are in much better agreement with the theory of Halpern and Hall than with that of Bethe and Bloch. The slight discrepancy below 1000 g cm^{-2} of water can be attributed to loss of particles by scattering out of the coverage of the counter telescope. Because of the directional selectivity of the telescope it is plausible to assume that at these shallow depths more mesons will be scattered away from the telescope than scattered in, resulting in a net loss of particles.

The discrepancy above 1000 g cm^{-2} portends the very large departure found by Rathgeber (1951) in a similar comparison extended down to very great depths.

The results for lead show very different behaviour. The departure from theory at about 1000 g cm^{-2} of lead is in the wrong direction to be explained by scattering. It will be seen from Figure 2 that the results obtained in a similar experiment by Heyland and Duncanson (1953*b*), when corrected for scattering, exhibit an even greater departure from the theoretical curve. The uncorrected data of these authors are found to coincide with the results of the present experiment.

VI. ACKNOWLEDGMENTS

The author is indebted to Professor L. H. Martin for his active interest in the problem and to Dr. H. D. Rathgeber for advice in the design of the experiment. Thanks are due to Mr. H. Waters and his associates who assisted in maintenance of the cosmic ray station. The cooperation of the authorities of the City of Box Hill is also acknowledged.

VII. REFERENCES

- BLACKETT, P. M. S. (1937).—*Proc. Roy. Soc. A* **159**: 2.
 BRINI, D., RIMONDO, O., and FILOSOFO, I. (1952).—*Nuovo Cim.* **9**: 505.
 CARO, D. E., PARRY, J. K., and RATHGEBER, H. D. (1951).—*Aust. J. Sci. Res. A* **4**: 16.
 DYER, A. J. (1953).—*Aust. J. Phys.* **6**: 61.
 EHMERT, A. (1937).—*Z. Physik* **106**: 751.
 GEORGE, E. P. (1952).—"Progress in Cosmic Ray Physics." (Ed. J. G. Wilson.) (Interscience Publishers: New York.)
 GLASER, D. A., HAMERMESH, B., and SAFONOV, G. (1950).—*Phys. Rev.* **80**: 625.
 HALPERN, O., and HALL, G. (1948).—*Phys. Rev.* **73**: 477.
 HAYAKAWA, S., and TOMONAGA, S. (1949).—*Progr. Theor. Phys. Osaka* **4**: 496.
 HEYLAND, G. R., and DUNCANSON, W. E. (1953*a*).—*Proc. Phys. Soc. Lond. A* **66**: 33.
 HEYLAND, G. R., and DUNCANSON, W. E. (1953*b*).—*Proc. Phys. Soc. Lond. A* **66**: 40.
 JANOSSY, L. (1948).—"Cosmic Rays." (Clarendon Press: Oxford.)
 MACKNIGHT, M. L., and CHASSON, R. L. (1951).—*Rev. Sci. Instrum.* **22**: 700.
 MONTGOMERY, D. J. X. (1947).—"Cosmic Ray Physics." (Princeton Univ. Press.)
 RATHGEBER, H. D. (1951).—*Z. Naturf.* **6a**: 578.
 ROSSI, B. (1948).—*Rev. Mod. Phys.* **20**: 537.
 WILSON, V. (1938).—*Phys. Rev.* **53**: 337.
 YORK, C. M. (1952).—*Phys. Rev.* **85**: 998.

THE LATTICE COMPONENT OF THE THERMAL CONDUCTIVITY OF METALS AND ALLOYS

By P. G. KLEMENS*

[*Manuscript received August 31, 1953*]

Summary

Makinson's (1938) theory of the lattice component of the thermal conductivity of metals and alloys, when limited at low temperatures by interaction with the conduction electrons, is re-examined, and the magnitude of the lattice conductivity is related to the electronic thermal conductivity at low temperatures, thus avoiding uncertainties in the theory at high temperatures. The result depends on whether transverse lattice waves can interact with the electrons.

The variation of the lattice conductivity with electron concentration is discussed, and the theory is applied to the systems copper-nickel, copper-zinc, silver-palladium, and silver-cadmium. At present only the first system has been measured, and the results can be reconciled qualitatively to the theory if it is assumed that holes appear in the $3d$ band for nickel concentrations greater than 10–20 per cent. It also appears that for copper there is direct interaction between electrons and transverse waves. Qualitative predictions are made for the other three systems.

I. INTRODUCTION

In addition to the thermal conduction in metals due to the free electrons, there is heat transport by the lattice. Since the free electrons, interacting with the lattice waves, limit the phonon mean free path, this lattice conductivity is much smaller than the conductivity of a dielectric solid and forms only a small part of the total thermal conductivity of a good electronic conductor. However, the lattice component becomes appreciable for metals and alloys of smaller electronic conductivity, and in some cases it has been possible to identify the lattice component.

The theory of the thermal conductivity of dielectric solids has been formulated by Peierls (1929) and extended by Klemens (1951). The same theory can be used in the presence of free electrons, except that the interaction between electrons and lattice waves must be considered as an additional scattering process. The relaxation time for this interaction has been worked out by Bethe (see Sommerfeld and Bethe 1933), and his result was used by Makinson (1938), who obtained an expression for the lattice conductivity in terms of electron interaction, boundary scattering, scattering by lattice defects, and phonon-phonon interaction and gave a qualitative discussion of the overall temperature dependence of the lattice conductivity.

* Division of Physics, C.S.I.R.O., University Grounds, Sydney.

In the special case when the only processes contributing significantly to the thermal resistance are electron interactions, Makinson obtained for the thermal conductivity

$$\kappa_g = 7.18G \left(\frac{T}{\theta} \right)^2, \quad \text{for } T \ll \theta, \quad \dots \dots \dots (1)$$

where it can be shown that

$$G = \frac{8\pi^2 \theta^2 K^3 M}{h^3 a^3 C_j^2} \left(\frac{1}{k} \frac{dE}{dk} \right)_\zeta^2 \dots \dots \dots (2)$$

Here θ is the Debye temperature, K the Boltzmann constant, M the mass, and a^3 the volume, of a unit cell, and E and k the energy and wave number of an electron state. The constant C_j is the usual coupling constant, of the dimensions of energy.

In order to eliminate the coupling constant C_j^2 , Makinson expressed κ_g in terms of $\kappa_e(\infty)$, the electronic thermal conductivity at high temperatures. He treated modes of all three polarizations as interacting equally with the conduction electrons, so that the constant C_j^2 in (2) is one-third of the corresponding constant appearing in the expression for electronic conductivity. With this assumption he obtained

$$\kappa_g = \frac{27}{4\pi^2 N^2} 7.18 \left(\frac{T}{\theta} \right)^2 \kappa_e(\infty), \quad \dots \dots \dots (3)$$

where N is the number of free electrons per atom. It should be emphasized that (3) depends on the assumption of a free electron gas, and also that $\kappa_e(\infty)$ is not the observed conductivity but the ideal conductivity (i.e. after subtracting the residual thermal resistance) of the same material on the assumption that the Bloch theory is obeyed at high temperatures.

Now it is well known that the Bloch theory neglects Umklapp-processes, which are certainly significant at high temperatures, that it disregards the dispersion of the velocity of sound, and that it does not allow for a possible frequency variation of the interaction constant C . Furthermore, the theory does not give the observed relation between the high temperature and the low temperature thermal conductivity. It seems thus more advantageous to eliminate the high temperature conductivity and to compare κ_g to the ideal electronic thermal conductivity at low temperatures. There still remains an uncertainty due to different possible coupling schemes. We shall discuss this point in Section II.

From (3) it appears at first sight that a variation of κ_g with the concentration of free electrons is accounted for by the factor N^2 in the denominator. It must be remembered, however, that $\kappa_e(\infty)$ itself varies with electron concentration. Unfortunately the ideal electronic thermal conductivity is generally unobservable for those substances whose lattice conductivity can be determined. The variation of κ_g with electron concentration must be determined directly from (2). This variation will be discussed in Section III. In Section IV the theory will be applied to the alloy systems copper-nickel, copper-zinc, silver-palladium, and silver-cadmium.

II. THE ABSOLUTE VALUE OF THE LATTICE CONDUCTIVITY

We shall express (2) in terms of the ideal electronic thermal conductivity $\kappa_e(T)$ at low temperatures. Sondheimer (1950) has obtained the following expression for the thermal conductivity at low temperatures according to the Bloch theory

$$\kappa_e(T) = \kappa_e(\infty) \left(\frac{\theta}{T} \right)^2 \frac{1}{71 \cdot 6 N^{2/3}}, \quad \dots \dots \dots (4)$$

solving the Bloch equation to third order by a variational method. The author (Klemens 1954) has solved the same equation numerically and obtained, with an uncertainty less than 0.5 per cent., a numerical factor 64.0 in place of 71.6. Using this new value and substituting into (3) we obtain

$$\kappa_g = 313 \kappa_e(T) \left(\frac{T}{\theta} \right)^4 N^{-4/3}. \quad \dots \dots \dots (5)$$

The factor $N^{-4/3}$ arises because κ_g is proportional to $k_{\zeta}^{-2} (dE/dk)_{\zeta}^2$, while κ_e is proportional to $k_{\zeta}^2 (dE/dk)_{\zeta}^2$. The factor $(dE/dk)_{\zeta}^2$ cancels in the ratio, so that in (5)

$$N^{4/3} \propto k_{\zeta}^4. \quad \dots \dots \dots (6)$$

This is independent of the dependence of E on k and, provided the Fermi surface is spherical, N denotes the actual number of electrons in the conduction band. If the Fermi surface touches the zone boundary, the effective value of $N^{2/3}$ will be reduced and will be proportional to the area of the Fermi surface in k -space; but for monovalent metals it is not likely to depart very much from the value $N=1$.

Formula (5) has been deduced on Makinson's assumption that electrons interact with the transverse as well as with the longitudinal waves, so that, if C^2 is the interaction constant in the expressions for the electronic conductivities, $C_j^2 = C^2/3$. In this case we must also take θ to be near the value θ_D , the Debye temperature as deduced from low temperature specific heat measurements.

However, if we adopt the Bloch model, in which electrons interact with the longitudinal lattice vibrations only, (5) must be modified. In this case $C_L^2 = C^2$, $C_T^2 = 0$. However, (1) and (2) will remain unaltered. The contribution to the thermal conductivity from the transverse phonons will not be higher than that given by Makinson's coupling model, if we assume that longitudinal and transverse phonons interact by means of three-phonon processes conserving total wave-vector. As has been pointed out elsewhere (Klemens 1951), these processes do not contribute directly to the thermal resistance, but are important in establishing thermal equilibrium and tend to equalize the effective mean free paths of phonons of the same frequency but different polarization.

In (3) we must thus replace C_j^2 by C^2 , so that an additional factor 1/3 enters into (5), the electronic conductivity being increased relative to the lattice conductivity. Thus

$$\kappa_g = 105 \kappa_e(T) \left(\frac{T}{\theta} \right)^4 N^{-4/3}, \quad \dots \dots \dots (7)$$

for interaction of electrons with longitudinal waves only. Also, in that case the appropriate θ -value will not be θ_D but θ_L , the Debye temperature appropriate to the longitudinal lattice waves only, as introduced by Blackman (1951).

Using values of the ideal thermal resistance observed by Berman and MacDonald (1952) for copper, and by White (1953*a*, 1953*b*) for silver and gold, the lattice thermal conductivity has been calculated from (5) using θ_D for θ , and again from (7), using θ_L . The value of θ_L for copper has been taken from Blackman (1951), and the values of silver and gold were chosen in the same proportion to θ_D . The results are shown in Table 1.

TABLE 1

Metal	$T^2\kappa_e(T)$		θ_D	θ_L	$\kappa_g T^{-2}$ according	$\kappa_g T^{-2}$ according
	(W cm ⁻¹ deg)		(°K)	(°K)	to Eqn. (5)	to Eqn. (7)
					(W cm ⁻¹ deg ⁻³)	(W cm ⁻¹ deg ⁻³)
Gold	7.7×10^3		170	270	2.9×10^{-3}	1.5×10^{-4}
Silver	$1.5_6 \times 10^4$		215	340	2.3×10^{-3}	1.2×10^{-4}
Copper	$4.4_5 \times 10^4$		315	505	1.4×10^{-3}	7.2×10^{-5}

Makinson's coupling scheme (interaction between electrons and waves of all polarizations) gives a lattice conductivity higher by a factor 20 than given by Bloch's coupling scheme (electrons interacting with longitudinal waves only), and observations of the lattice component of the thermal conductivity should thus provide a method of discriminating between these two schemes. It must be emphasized, however, that intermediate coupling schemes, such that $C_L^2 > C_T^2 \neq 0$, are also possible, and these would lead to intermediate values of the lattice conductivity. Finally, there is the possibility that the transverse waves are not coupled tightly to the longitudinal waves. This would result in an additional contribution to the conductivity, of different temperature dependence, analogous to the "longitudinal" conductivity of quartz glass (Klemens 1951).

While it is impossible to observe the lattice conductivity of the monovalent metals directly, it should be possible to infer it from the lattice conductivity of dilute alloys.

III. VARIATION OF LATTICE CONDUCTIVITY WITH ELECTRON CONCENTRATION

It is possible to vary the number of free electrons in a band by varying the composition of some alloys. The resulting variation of the lattice conductivity can serve thus as a qualitative test of Makinson's theory. Nevertheless, for this purpose one cannot use (3), because of the unknown behaviour of $\kappa_e(\infty)$, but (1) and (2) must be used directly. If the thermal resistance of the lattice component is $W_g = 1/\kappa_g$,

$$W_g \propto (\hbar dk/dE)\zeta^2, \dots\dots\dots (8)$$

for a spherical Fermi surface. This can be expressed also as

$$\begin{aligned} W_g &\propto g(k)\zeta/(dE/dk)\zeta^2 \\ &\propto G(E)\zeta/(dE/dk)\zeta, \dots\dots\dots (9) \end{aligned}$$

where $G(E)$ is the density of states per unit energy interval and $g(k)$ is the density of states per unit wave-number. The latter quantity is proportional to the area of the Fermi surface in k -space and it is easily seen from Sommerfeld and Bethe (1933), equation 46.2, that every element of the Fermi surface contributes additively to W_g , so that (9) is a general expression, valid also if the Fermi surface is non-spherical, or extends over two zones, provided dE/dk is averaged over all elements of the Fermi surface.

For a single band of free electrons, (9) is constant and independent of electron concentration. However, as the Fermi surface approaches the zone boundary, dE/dk decreases and W_g increases until the zone is nearly full, when $g(k)$ decreases and both $g(k)$ and W_g approach zero for a full band. In Figure 1 the variation of W_g with electron concentration for an s -band is shown schematically.

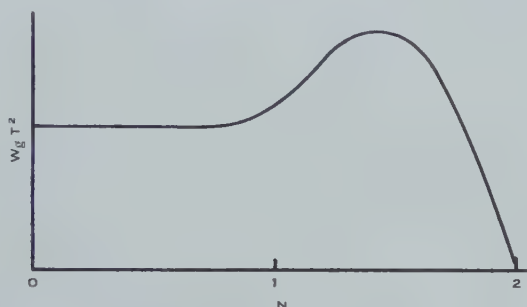


Fig. 1.—Variation of $W_g T^2$ with electron concentration for an s -band (schematic).

Such behaviour can be expected for the alloys copper-zinc and silver-cadmium. With increasing concentration of the divalent constituent the number of electrons in the s -band increases from one. A quantitative test seems impossible at present, because it is not known whether every divalent atom donates an additional electron to the conduction band or keeps a fraction of the additional electronic charge localized. Also the energy contours are not known in detail near the zone boundary and it is still an open question whether the Fermi surface of the monovalent metal touches the zone boundary.

In the case of the alloy systems copper-nickel and silver-palladium the number of electrons in the s -band is reduced with increasing concentration of the transition element. The thermal resistance due to the s -band should thus remain constant or decrease slightly. But at sufficiently high concentrations of the transition element, holes will appear in the d -band. It is seen from (9) that the resistance due to the d -band must be added to the resistance due to the s -band. The general variation of this component of W_g , at least for low concentrations of holes, will be given by a curve similar to the right-hand part of Figure 1. The total resistance will thus remain almost constant, and not more than the resistance of the monovalent metal, until that alloy composition at which holes first appear; then it will pass through a maximum and reach a constant value, above the value for the monovalent metal.

IV. COMPARISON WITH OBSERVATIONS

Apart from various technical alloys which are difficult to interpret, the only alloys whose thermal conductivity has been studied at low temperatures are the series copper-nickel, namely, 10 per cent. nickel by Estermann and Zimmerman (1952), 20 per cent. nickel by Hulm (1951), 30 per cent. nickel by Wilkinson and Wilks (1949), and 40 per cent. nickel by Berman (1951). Estermann and Zimmerman have also studied "Monel", an alloy containing 67 per cent. nickel, 30 per cent. copper, and iron, manganese, silicon, and carbon. We can regard "Monel" as approximately equivalent to 70 per cent. nickel and 30 per cent. copper.

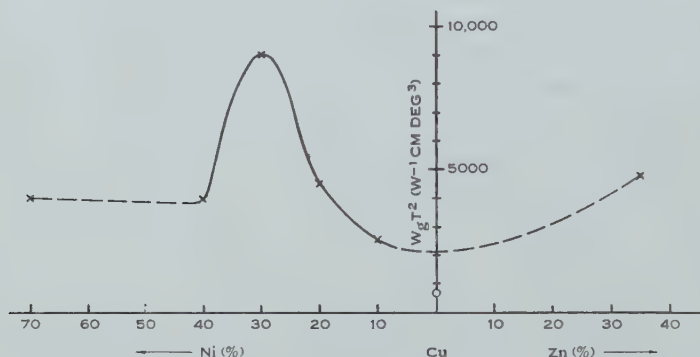


Fig. 2.—Variation of $W_g T^2$ with alloy composition for the series nickel-copper-zinc. Observed values \times , value expected for copper from equation (5) \circ . The value for copper-35 per cent. zinc is surmised from German silver (Berman 1951).

In Figure 2 the observed values of $W_g T^2$ are plotted against alloy composition, and the value calculated for pure copper according to Makinson's coupling scheme (equation (5)) is also shown. According to Bloch's coupling scheme (equation (7)), $W_g T^2$ is 20 times higher, and, unless there are unsuspected and violent variations between copper-10 per cent. nickel and pure copper, the experimental evidence seems to favour an intermediate coupling scheme, that is, electrons interacting with transverse phonons, but not as strongly as with longitudinal phonons.

An interesting feature of Figure 2 is the apparent appearance of an observable number of holes in the $3d$ -band at low nickel concentrations of about 15 per cent. This is in disagreement with the usual band theory, which predicts holes to first appear at 40 per cent. nickel (Mott 1935). On the other hand, there is other evidence for the anomalous appearance of holes at low nickel concentrations (Coles 1952), which confirms this interpretation.

The dotted line in Figure 2 indicates qualitatively the probable variation, inferred from the above considerations, of $W_g T^2$ with alloy composition for higher electron concentrations. It would be desirable to measure copper-nickel alloys of still lower nickel concentration (say 5 and 2 per cent.) and various copper-zinc alloys to test this behaviour. It may be that an examination of these alloys

will reveal a decrease of $W_g T^2$ below the value of 10 per cent. nickel. This would indicate better agreement with Makinson's coupling scheme and also the presence of holes for still lower nickel concentrations.

According to Coles the system silver-palladium behaves in agreement with the simple band theory. It would thus be of interest to study this system, as well as silver-cadmium, in order to test whether $W_g T^2$ is substantially constant from silver to silver-40 per cent. palladium (when holes should first appear) and whether $W_g T^2$ rises gradually with increasing concentration of cadmium for silver-cadmium, as expected. Such a study would clarify also the question of the interaction of transverse phonons with electrons, and possibly the departure from a spherical Fermi surface.

V. ACKNOWLEDGMENTS

The author wishes to acknowledge helpful criticism from Dr. R. E. B. Makinson, Mr. A. F. A. Harper, and Dr. G. K. White.

VI. REFERENCES

- BERMAN, R. (1951).—*Phil. Mag.* **42**: 642.
 BERMAN, R., and MACDONALD, D. K. C. (1952).—*Proc. Roy. Soc. A* **211**: 122.
 BLACKMAN, M. (1951).—*Proc. Phys. Soc. Lond. A* **64**: 681.
 COLES, B. R. (1952).—*Proc. Phys. Soc. Lond. B* **65**: 221.
 ESTERMANN, I., and ZIMMERMAN, J. E. (1952).—*J. Appl. Phys.* **23**: 578.
 HULM, J. K. (1951).—*Proc. Phys. Soc. Lond. B* **64**: 207.
 KLEMENS, P. G. (1951).—*Proc. Roy. Soc. A* **208**: 108.
 KLEMENS, P. G. (1954).—*Aust. J. Phys.* **7**: 64.
 MAKINSON, R. E. B. (1938).—*Proc. Camb. Phil. Soc.* **34**: 474.
 MOTT, N. F. (1935).—*Proc. Phys. Soc. Lond.* **47**: 571.
 PEIERLS, R. E. (1929).—*Ann. Phys. Lpz.* (5) **3**: 1055.
 SOMMERFELD, A., and BETHE, H. (1933). Elektronentheorie der Metalle. In "Handbuch der Physik." (Ed. H. Geiger and K. Scheel.) Vol. 24, Pt. 2. p. 333. (J. Springer: Berlin.)
 SONDHEIMER, E. H. (1950).—*Proc. Roy. Soc. A* **203**: 75.
 WHITE, G. K. (1953a).—*Proc. Phys. Soc. Lond. A* **66**: 559.
 WHITE, G. K. (1953b).—*Proc. Phys. Soc. Lond. A* **66**: 844.
 WILKINSON, K. R., and WILKS, J. (1949).—*J. Sci. Instrum.* **26**: 19.

THE THERMAL CONDUCTIVITY OF PURE METALS AT LOW TEMPERATURES ACCORDING TO THE FREE ELECTRON THEORY

By P. G. KLEMENS*

[*Manuscript received October 26, 1953*]

Summary

The paper discusses the validity of Sondheimer's (1950) variational method of solving the integral (Bloch) equation for the distribution of free electrons, interacting with lattice vibrations, in the case of thermal conduction at low temperatures. This equation is solved numerically, and it is found that the resulting thermal conductivity is about 11 per cent. larger than the value calculated by Sondheimer.

I. INTRODUCTION

It can easily be shown that the thermal conductivity of a pure metal, calculated on the assumption of energy transport by free electrons interacting with lattice vibrations but in the absence of scattering by static imperfections, should be inversely proportional to the square of the absolute temperature at sufficiently low temperatures. However, no analytical solution has been found to the Bloch integral equation which must be solved to obtain the multiplicative constant to that relationship.

The variational method of Kohler (1948, 1949), extended by Sondheimer (1950), overcomes the difficulty by adjusting the parameters of a linear trial function by a variational principle which leads to a stationary expression for the conductivity. The particular trial functions chosen are polynomials in the electron energy.

Clearly the results of such calculations must be sensitive, at least to some degree, to the particular form of the trial function used. Since it can be shown that the actual solution of the Bloch equation cannot be represented by a polynomial in electron energy, some doubt is thrown on the accuracy of Sondheimer's approximation. In order to clarify this point, the Bloch equation has been solved numerically for the region in which the T^{-2} law applies, where the error of the Sondheimer method is probably largest. The present result is estimated to be accurate to better than 0.5 per cent., and it is found that the theoretical conductivity exceeds Sondheimer's value by 11 per cent.

II. THE VARIATIONAL METHOD

Consider the equation

$$S.c(E) = E^n, \dots\dots\dots (1)$$

where $c(E)$ is a function of E , later to be identified as the electron energy, and S is a linear operator, so that $S.c(E)$ is also a function of E . Consider a real

* Division of Physics, C.S.I.R.O., University Grounds, Sydney.

Hilbert space, whose elements are all self-adjoint and are the functions c . Scalar products are defined by the integrals

$$(c_1, c_2) = (c_2, c_1) = \int c_1(E) c_2(E) \frac{df^0}{dE} dE, \quad \dots\dots\dots (2)$$

$f^0(E)$ being the Fermi distribution function. Let S have the special property

$$(Sc_1, c_2) = (Sc_2, c_1), \quad \dots\dots\dots (3)$$

and

$$(Sc_1, c_1) \geq 0. \quad \dots\dots\dots (4)$$

It is then easily shown by considering the variation of c that the solution of (1) is such that

$$(Sc, c) \text{ is a maximum, } \dots\dots\dots (5)$$

subject to the normalization condition

$$(Sc, c) = (E^n, c). \quad \dots\dots\dots (6)$$

In the Kohler-Sondheimer method, S is the integral operator describing the rate of change of $c(E)$ —a measure of the deviation of the distribution function from equilibrium—due to the interaction of the free electrons with the lattice vibrations. It was shown by Kohler (1948), that (3) and (4) are indeed satisfied.

Let $c^{(3/2)}$ and $c^{(5/2)}$ be the solutions of (1) with $n=3/2$ and $n=5/2$. Defining coefficients

$$K_{m,n} = K_{n,m} = (c^{(m+1/2)}, E^{n+1/2}) = (c^{(n+1/2)}, E^{m+1/2}), \quad \dots\dots\dots (7)$$

then the electrical conductivity is proportional to $K_{1,1}$, while the thermal conductivity is proportional to

$$\frac{K_{2,2}K_{1,1} - (K_{1,1})^2}{K_{1,1}T}. \quad \dots\dots\dots (8)$$

It is obvious from (5) and (6) that the coefficients $K_{2,2}$ and $K_{1,1}$ are stationary; that is, given a trial function deviating from the true solution by δc , the error in these coefficients will be of order $(\delta c, \delta c)$. The coefficient $K_{1,2}$ is not stationary for general variations, but Makinson (personal communication) has shown that, with the trial function a polynomial in E and its variations thus restricted, the Sondheimer method also leads to a stationary expression for $K_{1,2}$. The expression (8) for the thermal conductivity is then stationary.

The same result would be obtained if the Bloch equation for the case of thermal conduction were written in the form given by Bethe (see Sommerfeld and Bethe 1933), who showed that the thermal conductivity, correct to the first significant order in KT/ζ , is given by

$$\kappa \propto (c, \varepsilon), \quad \dots\dots\dots (9)$$

where $c(\varepsilon)$ is the solution of

$$S \cdot c(\varepsilon) = \varepsilon, \quad \dots\dots\dots (10)$$

and $\varepsilon = (E - \zeta)/KT$. The variational method can again be applied to (10). The expression (9) is of similar form to $K_{1,1}$ or $K_{2,2}$, and it is easily seen to be

stationary and a maximum. This formulation of the problem is more easily discussed than Kohler's. The function $c(\varepsilon)$ is, except for terms of higher order in KT/ζ , a linear combination of $c^{(3/2)}$ and $c^{(5/2)}$.

In the Kohler-Sondheimer method, the trial functions $c(E)$ are expressed in a series of ascending powers of E , whose coefficients are evaluated and substituted into the stationary expression for the transfer coefficients. This is equivalent to expressing $c(\varepsilon)$ as a power series in ε , and substituting the coefficients into (9). In practice only terms up to the cubic are retained. Now the function $c^{(3/2)}$ can be expressed very well by the first few terms of a power series; but the function $c^{(5/2)}$ cannot be thus represented at low temperatures for a pure metal, because the solution $c(\varepsilon)$ of (10)—given explicitly in (11) below—approaches a constant value asymptotically, as is easily seen by inspection. Thus the variational method, while giving good results for electrical conduction, and for thermal conduction when the Wiedemann-Franz law is obeyed, cannot give the correct solution $c(\varepsilon)$ for thermal conduction in a pure metal at low temperatures.

It does not necessarily follow that the variational method gives a value of the thermal conductivity seriously in error, for, if $c_p(\varepsilon)$ is the polynomial trial function, the approximation (c_p, ε) to (9) differs from the true value only by a term of second order in $\delta c = c_p - c$. However, this term is not necessarily small. In the real Hilbert space introduced above, all functions $c_p(\varepsilon)$ which are cubic polynomials form a subspace P . The true solution $c(\varepsilon)$ evidently lies outside this subspace, since it cannot be represented by a cubic. Thus the length $(\delta c, \delta c)^{1/2}$ cannot be less than that of the component of $c(\varepsilon)$ orthogonal to P , and this sets a lower limit to the second order error in (c, ε) . It may be that $(\delta c, \delta c)$ is not negligible compared with (c, c) .

It does not seem possible to give more than this qualitative discussion of the error, which from the above considerations must of course be negative. Equation (10) has therefore been solved numerically for a pure metal at very low temperatures, and the conductivity, varying as T^{-2} , has been evaluated from (9).

III. NUMERICAL SOLUTION OF THE BLOCH EQUATION

Disregarding multiplicative constants, the explicit form of (10) at low temperatures ($T \ll \theta$) in the absence of scattering by static imperfections is

$$\int_0^\infty \frac{x^2 dx}{e^x - 1} \left\{ [c(\varepsilon + x) - c(\varepsilon)] \frac{e^\varepsilon + 1}{e^\varepsilon + e^{-x}} + [c(\varepsilon - x) - c(\varepsilon)] \frac{e^{-\varepsilon} + 1}{e^{-\varepsilon} + e^{-x}} \right\} = \varepsilon. \quad \dots (11)$$

The symmetry of (11) requires that $c(\varepsilon)$ shall be an odd function of ε , and it is seen easily that $c(\varepsilon) \propto \varepsilon$ for small ε , and that $c'(\varepsilon) \propto \varepsilon^{-2}$ for large ε .

Two methods were used to solve (11) numerically, both of which gave substantially the same results. The lengthy computations involved were done on an automatic desk calculator.

In the first method the integral equation was replaced by a set of 10 simultaneous linear equations, with the values of $c(\varepsilon)$ at 10 equally spaced points as unknowns. The coefficients of the equations were the values of the kernel at these discrete points. Use was made of the property $c(-\varepsilon) = -c(\varepsilon)$, $c(0) = 0$.

It was assumed that $c(\varepsilon)=0$ for all values beyond the 10th point. The set of 10 equations was solved by a method of successive elimination adapted for desk machines, as described for example by Milne (1949). Two such calculations were made, using spacings of 0.5 and 1.0 respectively, and the two sets of solutions are marked in Figure 1. The assumption $c(\varepsilon)=0$ for large ε was seen to be inconsistent with the results, but, since the two sets of results did not differ greatly in spite of the different cut-off, it seems that the effect of this cut-off is not critical.

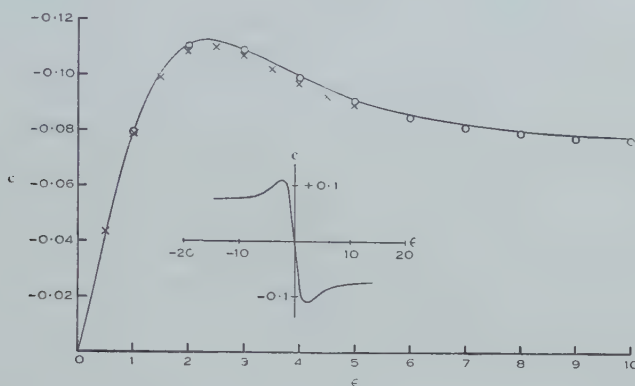


Fig. 1.—Solution of equation (10), replacing the integral equation by a discrete and finite set of linear equations. Crosses for $\varepsilon=0.5, 1, \dots, 5$; circles for $\varepsilon=1, 2, \dots, 10$. Full curve: result of the iterative method. Inset shows the form of $c(\varepsilon)$.

It would have been desirable to shorten the intervals without decreasing the range of unknown values. Unfortunately the methods of computation available did not permit a significant increase in the number of unknowns. However, it was noticed that in each equation the diagonal term was dominant, which suggested the following iterative procedure:

Let $c(\varepsilon)=c_0(\varepsilon)+c_1(\varepsilon)+c_2(\varepsilon)+\dots$, where $c_0(\varepsilon)$ is given by

$$\int_0^\infty \frac{x^2 dx}{e^x - 1} \left[\frac{e^\varepsilon + 1}{e^\varepsilon + e^{-x}} + \frac{e^{-\varepsilon} + 1}{e^{-\varepsilon} + e^{-x}} \right] c_0(\varepsilon) = -\varepsilon, \dots \dots \dots (12)$$

and

$$\int_0^\infty \frac{x^2 dx}{e^x - 1} \left\{ [c_{n-1}(\varepsilon+x) - c_n(\varepsilon)] \frac{e^\varepsilon + 1}{e^\varepsilon + e^{-x}} + [c_{n-1}(\varepsilon-x) - c_n(\varepsilon)] \frac{e^{-\varepsilon} + 1}{e^{-\varepsilon} + e^{-x}} \right\} = 0. \dots \dots \dots (13)$$

Evaluating $c_0(\varepsilon)$, $c_1(\varepsilon)$, $c_2(\varepsilon)$, etc. in turn from these definite integrals, successive approximations were obtained. The integrations were done numerically and $c_{n-1}(\varepsilon)$ was obtained for a few values of ε . The intermediate values were obtained by interpolation and used for the next iteration. The iteration was broken off after c_3 . The smallness of $c_n(\varepsilon)$ indicates whether $c(\varepsilon)=c_0(\varepsilon)+c_1(\varepsilon)+\dots+c_{n-1}(\varepsilon)$ is a good approximation. While the convergence of this method was not examined, the actual results, plotted in Figure 2, indicate that it is satisfactory.

No investigation was made of the accumulation of errors, and therefore the solution had to be tested by substitution into (11). The conductivity can then be obtained from (9). These two steps were combined and a further correction to the conductivity was obtained by multiplying the solution by an arbitrary constant and renormalizing it using the relation

$$(Sc, e) = (\epsilon, e). \quad (14)$$

As before, with e thus normalized, the expression (ϵ, e) is stationary. The renormalization correction was only 0.4 per cent.; hence the conductivity thus calculated is accurate to at least the same limit.

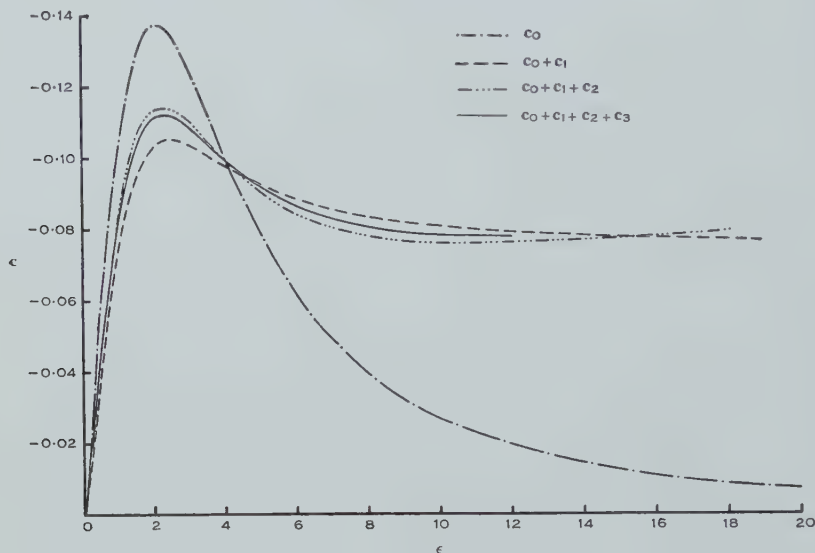


Fig. 2.—Solution of equation (9) by iteration.

IV. CONCLUSION

Including the multiplicative factors in the Bloch equation and the expression for the thermal conductivity, which have not been stated here, but are given by Bethe (see Sommerfeld and Bethe 1933, Section 39), the thermal conductivity in the T^{-2} region has been obtained from the numerical solution of (11). Expressing this in terms of a hypothetical thermal conductivity at high temperatures, assuming a Debye spectrum of limiting frequency $2\pi K\theta/h$, a spherical Fermi surface due to N free electrons per atom, and absence of any processes other than those considered in the Bloch theory for a pure metal (in particular no Umklapp-processes), one obtains

$$\kappa(T) = \kappa(\infty)(\theta/T)^2(64 \cdot 0 N^{2/3})^{-1}. \quad (15)$$

The result of Sondheimer's (1950) third approximation, when similarly expressed, has a numerical constant of 71.6 in place of 64.0, while the first approximation, obtained by Wilson (1937), leads to 95.3. Sondheimer's value differs from the present solution by only 11 per cent. As is to be expected, the present method gives a larger value for the conductivity than Sondheimer's.

In addition to the conductivity, these calculations also give the functional dependence of $c(\epsilon)$. The variational method, using a polynomial trial function, does not even approximate the true solution, though it gives reasonable values for the conductivity. While the function $c(\epsilon)$ is only of secondary interest, knowledge of it may be useful in other work.

There are well-known discrepancies between the Bloch theory and the thermal conductivity observed at low temperatures (Hulm 1950, 1952 ; Andrews, Webber, and Spohr 1951 ; Berman and MacDonald 1951, 1952 ; White 1953), which are hardly reduced by the result of the present calculations. It thus appears that the simple free electron model is inadequate. The required modifications of the model will be discussed subsequently.

V. ACKNOWLEDGMENTS

The author wishes to thank Dr. R. E. B. Makinson for valuable advice, and Miss J. Ward for performing the computations.

VI. REFERENCES

- ANDREWS, F. A., WEBBER, R. J., and SPOHR, D. A. (1951).—*Phys. Rev.* **84** : 994.
 BERMAN, R., and MACDONALD, D. K. C. (1951).—*Proc. Roy. Soc. A* **209** : 386.
 BERMAN, R., and MACDONALD, D. K. C. (1952).—*Proc. Roy. Soc. A* **211** : 122.
 HULM, J. K. (1950).—*Proc. Roy. Soc. A* **204** : 98.
 HULM, J. K. (1952).—*Proc. Phys. Soc. Lond. A* **65** : 227.
 KOHLER, M. (1948).—*Z. Phys.* **124** : 772.
 KOHLER, M. (1949).—*Z. Phys.* **125** : 679.
 MILNE, W. E. (1949).—"Numerical Calculus." (Princeton Univ. Press.)
 SOMMERFELD, A., and BETHE, H. (1933).—Elektronentheorie der Metalle. In "Handbuch der Physik." (Ed. H. Geiger and K. Scheel.) Vol. 24, Pt. 2. p. 333. (Springer : Berlin.)
 SONDHEIMER, E. H. (1950).—*Proc. Roy. Soc. A* **203** : 75.
 WHITE, G. K. (1953).—*Proc. Phys. Soc. Lond. A* **66** : 559, 844.
 WILSON, A. H. (1937).—*Proc. Camb. Phil. Soc.* **33** : 371.

THE ELECTRICAL AND THERMAL CONDUCTIVITIES OF MONOVALENT METALS

By P. G. KLEMENS*

[*Manuscript received November 5, 1953*]

Summary

The experimentally determined values of the high and low temperature electrical and thermal conductivities of pure sodium, copper, silver, and gold are such that the ratios of these quantities for each of these metals do not agree with the values expected from the Bloch free electron theory, except for the high temperature Wiedemann-Franz ratio. Reasonable agreement can be achieved by assuming (i) that the conduction electrons can interact directly with transverse lattice vibrations, and (ii) that the Fermi surface departs significantly from spherical shape in all these metals, and touches the boundary of the Brillouin zone in the case of copper, silver, and gold.

I. INTRODUCTION

The thermal conductivity of a pure metal at low temperatures, assuming energy transport by quasi-free electrons (electron energy E a function of the wave number k only) interacting with lattice vibrations having a Debye spectrum, has been calculated by the author (Klemens 1954b) by solving numerically the appropriate transport equation. The conductivity thus obtained is about 11 per cent. larger than the value previously obtained by Sondheimer (1950), and its accuracy was estimated to be better than 0.5 per cent. This does not substantially alter the well-known discrepancy between the predictions of the theory and the observed thermal conductivity. In this paper the discrepancy will be re-examined and modifications of the quasi-free electron model which offer an explanation will be considered.

In all theoretical expressions for the thermal and electrical conductivities of pure metals there enter two constants, always in the same combination, which are difficult to calculate but can be combined and treated as an empirical constant. They are the effective number of free electrons and the electron-phonon interaction constant C . When comparing experimental results with theoretical predictions, the empirical constant is eliminated by forming the ratio of two conductivities. Of the four quantities used, namely, the electrical and thermal conductivities at high and at low temperatures, the high temperature electrical and thermal conductivities are related by the Wiedemann-Franz law, so that only two independent ratios can be formed. The Wiedemann-Franz law holds at high temperatures irrespective of particular assumptions about the band structure and the electron-phonon interaction, and it is therefore not a sufficient test of the Bloch theory. It is in fact well obeyed at high temperatures

* Division of Physics, C.S.I.R.O., University Grounds, Sydney.

for all metals, except where there is an appreciable lattice component of thermal conduction.

By the quasi-free electron theory if W is the thermal resistance and T_1 and T_2 are two temperatures such that $T_1 \ll \theta < T_2$ then, according to Klemens (1954b),

$$\frac{W(T_1)}{W(T_2)} = 64 \cdot 0 N^{2/3} (T_1/\theta)^2, \quad \dots\dots\dots (1)$$

where θ is the Debye temperature and N the number of free electrons per atom. Also the electrical resistance at high and at low temperatures is related by

$$\frac{R(T_1)}{R(T_2)} = 497 \cdot 6 \frac{T_1^5}{T_2 \theta^4}, \quad \dots\dots\dots (2)$$

and, since $R(T_2) = L W(T_2) T_2$, where $L = (\pi K/e)^2/3 = 2 \cdot 45 \times 10^{-8} \text{ W}\Omega/\text{deg}^2$, the high temperature conductivities can be eliminated from (1) and (2) giving

$$\frac{W(T_1)}{T_1^2} = \frac{64 \cdot 0}{497 \cdot 6} N^{2/3} \frac{\theta^2}{L} \frac{R(T_1')}{T_1'^5}, \quad \dots\dots\dots (3)$$

where T_1 and $T_1' \ll \theta$. This equation could, of course, be derived without reference to the high temperature conductivities.

It has previously been assumed by various authors (see Sondheimer 1952) that N is the *effective* number of free electrons per atom, defined in terms of the current induced in the band by an electric field, and that it can be treated as an adjustable parameter. However, it arises in (1) from a term $q^2/2k\zeta$ in the transport equation at high temperatures, q being the maximum wave number of the phonons and $k\zeta$ the Fermi wave number. For a spherical Fermi surface $q/k\zeta \propto N^{1/3}$, where N is the number of free electrons per atom in the conduction band, quite independently of the dependence of E on k . For a monovalent metal we must therefore take $N=1$.

II. THERMAL CONDUCTIVITIES AT HIGH AND AT LOW TEMPERATURES

The thermal conductivities of pure sodium and copper have been measured by Berman and MacDonald (1951, 1952) and that of pure gold and silver by White (1953*a*, 1953*b*), who also repeated the measurements on copper (White 1953*c*). Subtracting the residual resistance, they obtained the ideal thermal resistance, which at low temperatures is indeed approximately proportional to T^2 . The values of $W(T_1)/T_1^2$ so obtained, together with $W(T_2)$, the thermal resistance at room temperatures, are given in Table 1. Taking for θ the value θ_R , which is derived by fitting the ideal electrical resistance to (2), and putting $N=1$, the ratio $W(T_1)/T_1^2 W(T_2)$ was calculated from (1) and compared with the corresponding ratio of the measured values. As seen in the table, these two ratios are not equal, so that (1) is not satisfied.

It has not been made clear whether this discrepancy arises from a failure of the theory in respect of $W(T_1)$ or of $W(T_2)$, or because of a wrong choice of the θ -value. The last two effects are not independent, for a failure of the theory to give the correct expression for $W(T_2)$ —and hence for $R(T_2)$ —would affect the

value of θ_R . The theoretical expressions for $W(T_2)$ and $R(T_2)$ may not be correct because the Bloch theory neglects the following effects which should be considered at high temperatures even for the quasi-free electron model:

(i) A lowering of the effective vibration frequencies due to the dispersion of the lattice waves, and due to each high frequency wave being neither purely longitudinal nor purely transverse.

(ii) The occurrence of Umklapp-processes. These are processes which conserve energy but change the total wave-vector by an inverse lattice vector. On the quasi-free electron model these processes cannot occur at low temperatures, but must occur at high temperatures. Since they are not considered in

TABLE I

Metal		Sodium	Copper	Silver	Gold
$W(T_1)/T_1^2$	(W^{-1} cm deg $^{-1}$)	3.8×10^{-4}	2.55×10^{-5}	6.4×10^{-5}	1.3×10^{-4}
$W(T_2)$	(W^{-1} cm deg) ..	0.73	0.26	0.24	0.64
$R(T_1)/T_1^5$	(Ω cm deg $^{-5}$) ..	$5.3_7 \times 10^{-15}$	$2.6_4 \times 10^{-16}$	1.11×10^{-15}	3.9×10^{-15}
θ_R (°K)	202	330	220	170
θ_D (°K)	150	315	215	170
θ_L (°K)	260	505	340*	270*
$\theta_{(3)}$ (°K)	117	136	105	80
Ratio of calculated to observed values of $W(T_1)/T_1^2 W(T_2)$					
using θ_R	3.0	6.0	5.0	5.8
θ_L	1.8	2.6	2.1	2.3
$(\theta_L/\theta_R)^4$	2.8	5.4	5.8	6.3
$(\theta_L/\theta_{(3)})^2$	5.0	13.7	10.5	11.4
$(\theta_D/\theta_{(3)})^2$	1.7	5.4	4.2	4.5

* Taking the same ratio of θ_L/θ_D as for copper.

the Bloch theory, they form an additional source of resistance, also proportional to T at high temperatures. The magnitude of this resistance, which is difficult to calculate accurately, is sensitive to the shape of the zone boundary and to its position relative to the Fermi surface. However, it can never greatly exceed the Bloch resistance, and would only be about 30 per cent. of the Bloch resistance for $N=1$ and a spherical Brillouin zone.

(iii) A possible variation at high frequencies of the electron-phonon interaction parameter C , assumed in the Bloch theory to be constant for all interaction processes. Such a variation would follow from Nordheim's assumption of a rigid ion, in contrast to Bloch's deformable ion, and would also occur in intermediate cases. But even for a deformable ion a decrease of C with increasing phonon frequency would be obtained if proper account were taken of the partial interference of electron wave-functions in calculating the transition matrix. This point is discussed by Bethe (see Sommerfeld and Bethe 1933, p. 517).

While (i) and (ii) would cause the high temperature resistance to increase over the value given by the Bloch theory, (iii) would cause a decrease.

In view of these uncertainties, equation (1) is not a good test of the validity of the theory of thermal conduction at low temperatures.

III. THERMAL AND ELECTRICAL CONDUCTIVITIES AT LOW TEMPERATURES

The uncertainties of the theory at high temperatures can be avoided by comparing the thermal and electrical conductivities at low temperatures, using equation (3). One can confidently expect the phonon spectrum at low frequencies to be of the form given by the simple elastic theory, so that, according to the Bloch theory, the θ -value in (3) should be θ_L , related to the velocity of longitudinal low frequency waves and calculated by Blackman (1951) for some metals, including sodium and copper.

Substituting the observed values of W/T^2 and R/T^5 into (3) and calculating a value for θ from it, denoted by $\theta_{(3)}$, we find that $\theta_{(3)}$ is too low in all cases (see Table 1). The discrepancy $(\theta_L/\theta_{(3)})^2$ is somewhat less for sodium than for the noble metals.

It should be noted that the discrepancy in (1) would have been reduced if θ_L had been used instead of θ_R ; but this could have been done only at the expense of introducing a discrepancy $(\theta_L/\theta_R)^4$ in (2) (Blackman 1951). Attempts to resolve this discrepancy, by Klemens (1952) in terms of dispersion of the lattice waves, and by Bhatia (1952) in terms of Umklapp-processes, are both in error.* A full consideration of these effects would lead to an increase in the high temperature resistance, as pointed out above, so that the discrepancy in (2) would be increased even further.

There are thus discrepancies between theory and experiment which can be explained neither in terms of deviations of the phonon spectrum from the Debye model, nor by reasonable adjustment of the θ -values. One is therefore led to conclude that they arise as result of deviations of the electronic band structure from the quasi-free electron model.

IV. MODIFICATIONS OF THE QUASI-FREE ELECTRON MODEL

It has been seen that the observed ratio of the electrical to the thermal resistance at low temperatures is greater than expected from the Bloch theory. The following explanation is offered.

It is well known that the processes responsible for electrical resistance are such as to move an electron in momentum space from a point on the Fermi surface to a point on the opposite side of it. At high temperatures this is done in large steps, each phonon interaction changing the direction of the electron by about 1.5 radians. At low temperatures the angular change at each interaction is only of order T/θ , so that the processes producing the electrical resistance

* Klemens takes the density of normal modes to be proportional to $\omega^2 d\omega$, ω being the phonon frequency, but this is correct only in the absence of dispersion. Bhatia correctly deduces that the θ -value at high temperatures is lowered because of Umklapp-processes, but does not consider that these processes increase $R(T_2)$, and hence increase the θ -value deduced from (2).

can be regarded as a small-step diffusion process, in which an electron wanders to the opposite side of the Fermi surface. The calculation of the resistance becomes a random walk problem on the Fermi surface. The electrical resistance is inversely proportional to the square of the distance to be covered and proportional to the number of steps per unit time and to the square of the average length of each step. It would therefore be expected that the electrical resistance of a metal having a non-spherical Fermi surface would differ from that derived on the Bloch theory, where a spherical Fermi surface is assumed.

On the other hand, the thermal resistance at low temperatures is due to a movement of electrons from a point just above the Fermi surface to one just below it, or vice versa; that is, a change of energy of order KT without an appreciable change of direction. Since in such processes the electron does not change its location on the Fermi surface but only its "height" above or below it, changes in the shape of the Fermi surface will not affect the thermal resistance, except by changing the effective number of free electrons, which does not enter the present considerations.

The possible deformations of the Fermi surface from spherical shape satisfy the following requirements: (i) the enclosed volume is kept constant, (ii) the deformation has the polyhedral symmetry of the Brillouin zone, and (iii) along the axes of symmetry the deformation is outward. Such a deformation will alter the electrical resistance relative to the thermal resistance at low temperatures. This effect will be even more pronounced, and increase the electrical resistance, if the Fermi surface touches the zone boundary, for then an electron, diffusing on the Fermi surface, can reach an opposite point not only by the usual way but also by drifting to the nearest point of contact and reappearing on the opposite side of the zone. The distance to be covered is thus approximately halved on the average, and there will be an additional resistance, about four times the ordinary resistance, due to movement via the points of contact, so that the total resistance is increased by a factor of about 5.

We have seen that the ratio $R(T_1)/W(T_1)$ is larger than expected from the Bloch theory, the discrepancy being $(\theta_L/\theta_{(3)})^2$ shown in Table 1. This discrepancy can be reduced by a factor of about 5 if it is assumed that the Fermi surface touches the zone boundary. Even so, for the noble metals, the discrepancy is too large to be explained in this way alone. In order to reconcile the theory to the observed values of $R(T_1)/W(T_1)$, it must be further assumed that the conduction electrons interact with the transverse waves as well as with the longitudinal waves. This would make θ_D the appropriate θ -value in equation (3) and the discrepancy in (3), now $(\theta_D/\theta_{(3)})^2$, is sufficiently reduced, as seen from Table 1, to permit an explanation in terms of the Fermi surface touching the zone boundary.

Peierls (1930*a*, 1930*b*, 1932) suggested that the Fermi surface of monovalent metals should touch the zone boundary in order to account for the absence of observable effects at low temperatures arising from quasi-equilibria between electrons and phonons, and, although Klemens (1951) showed that this was not a necessary conclusion, it remained a possibility. Smit (1952) pointed out that the effects of shear strain on the thermoelectric forces suggest strongly that the

Fermi surface touches the zone boundary in the cases of gold and silver. Measurements by Mortlock (1953) have made it appear probable that this is also the case for copper. The sign of the thermoelectric power of the noble metals, in contrast to that of the alkali metals, gives additional support to the hypothesis.

Bloch's conclusion that the conduction electrons cannot interact directly with transverse phonons is based on the assumption of a spherical Fermi surface. Since this assumption has here been discarded to explain the low temperature conductivities, it is not unreasonable to assume some interaction between the conduction electrons and transverse waves. There is also some evidence for such interaction from the study of the lattice component of the thermal conductivity of alloys (Klemens 1954a). The magnitude of the lattice component depends sensitively upon the degree of interaction between the conduction electrons and transverse waves. The lattice component of a pure metal cannot be determined directly but can be deduced from the thermal conductivity of dilute alloys. Measurements on copper-nickel alloys, as discussed by the author, indicate such interaction in the case of copper, and recent measurements of silver-palladium alloys (White, personal communication) indicate the same for silver.

Considering now the high temperature resistance (electrical or thermal), and remembering that for Fermi surfaces touching the zone boundary the contribution from Umklapp-processes to the resistance is roughly equal to that from ordinary processes, and also that dispersion can reduce the frequency of the shortest lattice waves by about 1.5 (Klemens 1952) and thus increase the resistance by a factor of 2 to 2.5, one can explain the discrepancies in (1) with $\theta = \theta_D \sim \theta_R$ for the noble metals. This would, of course, also explain the discrepancies in (2) noted by Blackman (1951).

With sodium the position is different, since the discrepancy from (3) is smaller. It can be explained either by assuming that the Fermi surface touches the zone boundary, but that the conduction electrons do not interact with transverse waves, or by assuming a Fermi surface, non-spherical but nevertheless not touching the zone boundary, and interaction between the electrons and transverse phonons. The latter explanation seems more probable, because it will also explain the observed value of the high temperature resistance, and because of the normal sign of the thermoelectric power.

V. CONCLUSIONS

The experimental value of the ratio R/W for pure metals at low temperatures is too large compared with the predictions of the Bloch theory, and deviations from the quasi-free electron model must be assumed to explain this discrepancy. Values of the high temperature resistance also indicate that the Bloch theory is not valid. It seems that in the four metals considered there is interaction between the conduction electrons and transverse lattice waves. In gold, silver, and copper the Fermi surface apparently touches the zone boundary, while in sodium it does not.

There is evidence in support of these conclusions from thermoelectric effects and from the lattice component of the thermal conductivity.

VI. ACKNOWLEDGMENTS

The author wishes to thank Dr. R. E. B. Makinson, Mr. A. F. A. Harper, and Dr. G. K. White for helpful suggestions and discussions.

VII. REFERENCES

- BERMAN, R., and MACDONALD, D. K. C. (1951).—*Proc. Roy. Soc. A* **209** : 368.
BERMAN, R., and MACDONALD, D. K. C. (1952).—*Proc. Roy. Soc. A* **211** : 122.
BHATIA, A. B. (1952).—*Proc. Phys. Soc. Lond. A* **65** : 188.
BLACKMAN, M. (1951).—*Proc. Phys. Soc. Lond. A* **64** : 681.
KLEMENS, P. G. (1951).—*Proc. Phys. Soc. Lond. A* **64** : 1030.
KLEMENS, P. G. (1952).—*Proc. Phys. Soc. Lond. A* **65** : 71.
KLEMENS, P. G. (1954a).—*Aust. J. Phys.* **7** : 57.
KLEMENS, P. G. (1954b).—*Aust. J. Phys.* **7** : 64.
MORTLOCK, A. J. (1953).—*Aust. J. Phys.* **6** : 410.
PIERLS, R. E. (1930a).—*Ann. Phys. Lpz.* **4** : 121.
PIERLS, R. E. (1930b).—*Ann. Phys. Lpz.* **5** : 244.
PIERLS, R. E. (1932).—*Ann. Phys. Lpz.* **12** : 154.
SMIT, J. (1952).—*Physica* **18** : 587.
SOMMERFELD, A., and BETHE, H. (1933).—Elektronentheorie der Metalle. In "Handbuch der Physik." (Ed. H. Geiger and K. Scheel.) Vol. 24, Pt. 2. p. 333. (Springer : Berlin.)
SONDHEIMER, E. H. (1950).—*Proc. Roy. Soc. A* **203** : 75.
SONDHEIMER, E. H. (1952).—*Proc. Phys. Soc. A* **65** : 562.
WHITE, G. K. (1953a).—*Proc. Phys. Soc. A* **66** : 559.
WHITE, G. K. (1953b).—*Proc. Phys. Soc. A* **66** : 844.
WHITE, G. K. (1953c).—*Aust. J. Phys.* **6** : 397.

X-RAY LINE BROADENING AND PURE DIFFRACTION CONTOURS

By R. I. GARROD,* J. F. BRETT,* and J. A. MACDONALD*

[*Manuscript received October 19, 1953*]

Summary

In analysing the data from experiments designed to distinguish between particle size and distortion broadening from polycrystalline materials, it is customary either to employ correction formulae to obtain the true broadening β , or to derive the pure diffraction contour in terms of a Fourier series whose coefficients may be evaluated from the experimental line profiles. The first method leads to values of β that are critically dependent upon the particular functions chosen to represent the diffraction line profiles and the second method, whilst removing this ambiguity, only yields the pure diffraction contour numerically and not analytically.

By applying Fourier methods, it is shown that the pure diffraction contours associated with particular causes of broadening can in fact be identified with certain types of analytic functions. In particular, the Cauchy and Gaussian distributions, which have often been arbitrarily employed in the past to represent the pure diffraction contour and experimental line profiles, are only strictly applicable to particular types of particle size and lattice distortion effects respectively. The case of combined size and distortion broadening is also considered, and for pure particle size broadening correction curves are derived corresponding to different types of particle size distributions.

I. INTRODUCTION

It is well known that radial broadening of X-ray diffraction lines from polycrystalline materials is associated with small particle size or variations in lattice spacing (e.g. due to faults in the crystal or heterogeneous lattice strains) over the volume of the material irradiated by the X-ray beam. In some cases it is clear which of these factors is predominant, but in others either effect, or a combination of them, may equally well be the cause of the broadening, and several methods have been suggested for differentiating between the various possibilities.

The problem is complicated by the necessity for correcting the broadening actually observed for instrumental effects before the true broadening due to the inherent condition of the material can be obtained. To carry out this correction it is customary either to employ correction formulae relating the true (required) integral line breadth β to the total observed breadth B and the instrumental breadth b (Scherrer 1920 ; Jones 1938 ; Warren and Biscoe 1938 ; Taylor 1941 ; Schoening, van Niekerk, and Haul 1952), or to analyse the experimentally determined line profiles by somewhat laborious mathematical procedures (Stokes 1948 ; Paterson 1950).

* Defence Research Laboratories, Melbourne.

In the first class, the values obtained for β , for given values of b/B , depend markedly, for the range covered by most experimental data, upon the particular correction formula adopted. Shull (1946) has shown that the true line profile after correcting for instrumental broadening (hereafter termed the "pure diffraction contour"), and hence the most appropriate correction formula, may be derived if the experimental line shapes are assumed or known to conform to particular types of analytic functions. Unfortunately, however, unless extremely accurate experimental techniques are employed, this criterion is of limited use, because the various types of functions which may be chosen to fit the observed profiles only differ appreciably in shape in regions close to their "tails" where possible percentage errors in measurement of line intensity are greatest. This uncertainty as to the proper relationship between β , B , and b is of particular consequence in attempting to distinguish between particle size and distortion broadening from deformed metals by investigating the variation of β with diffraction angle and X-ray wavelength, and has contributed to the conflicting nature of the results obtained in this field in the past (Brindley 1940; Smith and Stickley 1943; Stokes, Pascoe, and Lipson 1943; Megaw and Stokes 1945; Wood and Rachinger 1949).

Uncertainties of this type may be avoided by the use of the second class of method mentioned above and the more recent work on cold-worked metals has tended to adopt this type of procedure (Patterson 1950; Warren and Averbach 1950, 1952*a*, 1952*b*; Auld and Garrod 1952; McKeehan and Warren 1953; Smith 1953; Williamson and Hall 1953). In this case the pure diffraction contour is obtained numerically but not analytically.

Line breadth, however, is only one of the parameters associated with the broadened X-ray diffraction lines. Recently, it has been shown that much additional information may be obtained from a study of the shape of the pure diffraction contour. The latter is represented by a Fourier series obtained numerically from the experimental X-ray data by Stokes's method (1948), and the dependence of the Fourier coefficients on frequency can be used to investigate particle size broadening (Bertaut 1950, 1952; Warren and Averbach 1950), distortion broadening (Warren and Averbach 1950, 1952*a*) or a combination of the two effects (Eastabrook and Wilson 1952; Paterson 1952; Warren and Averbach 1952*b*).

It is important to note that in all of this work mentioned above, either analytic functions have been employed as direct or tacit assumptions for representing the X-ray line profiles or, alternatively, the pure diffraction contours have been obtained *numerically* from experimental data by computation. The question therefore arises whether in fact pure diffraction contours, due to particular causes of broadening, can be identified with particular types of analytic functions and, if so, the range of validity associated with the use of such functions.

In the present paper an attempt is made to investigate some aspects of this problem by the use of the Fourier methods referred to previously.

II. VALIDITY OF ANALYTIC FUNCTIONS FOR THE PURE DIFFRACTION CONTOUR

The following analysis for a randomly oriented polycrystalline material applies to any reflection which can be of the form $00l$ with respect to an appropriate system of orthogonal axes. As shown previously (Eastabrook and Wilson 1952; Paterson 1952), the pure diffraction contour for a material exhibiting line broadening may be represented quite generally by

$$I(X) = R \int_{-\infty}^{\infty} A(m, l) \exp(2\pi i m X) dm, \quad \dots\dots\dots (1)$$

where $X = (2\theta - 2\theta_0)\lambda^{-1} a_3 \cos \theta_0$,

a_3 = length of unit cell axis in $[00l]$ direction,

θ_0 = Bragg angle for peak intensity,

m = (variable) difference in coordinates in $[00l]$ direction between any pair of cells in the crystal,

R = constant for a particular experimental arrangement.

$A(m, l)$ is a Fourier transform of $I(X)$ and may be expressed as the product of two other quantities (Eastabrook and Wilson 1952),

$$A(m, l) = N(m)J(m, l), \quad \dots\dots\dots (2)$$

where $N(m)$ depends only upon the size and shape of the crystallites in the sample and $J(m, l)$ depends upon the lattice distortion.

If the analysis is restricted to diffraction contours that are symmetrical about the peak value, it follows that for any given $00l$ reflection, $A(m, l)$ may be written as

$$A(m) = P \int_{-\infty}^{\infty} I(x) \cos(2\pi m x / T) dx, \quad \dots\dots\dots (3)$$

where $x = 2(\theta - \theta_0)$,

$T = \lambda / (a_3 \cos \theta_0)$,

P = constant.

Eastabrook and Wilson (1952) have shown that, if $A(m)$ is plotted against m , the initial slope is a measure of the reciprocal of the mean particle size \bar{M} (in units of a_3) in the $[00l]$ direction and the initial curvature gives a lower limit to the mean square strain \bar{e}^2 ; that is,

$$-\left(\frac{d\hat{A}}{d|m|}\right)_{|m|=0} = (\bar{M})^{-1}, \quad \dots\dots\dots (4)$$

$$\left(\frac{d^2\hat{A}}{d|m|^2}\right)_{|m|=0} \simeq -4\pi^2 l^2 \bar{e}^2, \quad \dots\dots\dots (5)$$

where $\hat{A}(m) \equiv A(m)/A(0)$.

(a) *The Functions* $(1+a^2x^2)^{-1}$, $\exp(-k^2x^2)$, and $(1+c^2x^2)^{-2}$

The pure diffraction contour $I(x)$ is related to the intensity distribution $g(x)$ due to instrumental factors and the total distribution $h(x)$ from a material exhibiting line broadening by the equation (Jones 1938)

$$h(x) = \frac{\int_{-\infty}^{\infty} g(u)I(x-u)du}{\int_{-\infty}^{\infty} I(x)dx}, \quad \dots\dots\dots (6)$$

where u is the parameter of integration.

$h(x)$ and $g(x)$ are the experimentally observed line profiles, and various authors have from time to time suggested different analytic functions to represent the experimental data. Among these are the functions $(1+a^2x^2)^{-1}$, $\exp(-k^2x^2)$, and $(1+c^2x^2)^{-2}$.

It can easily be shown (Shull 1946) that, if $h(x)$ and $g(x)$ are of the form $(1+a^2x^2)^{-1}$, then $I(x)$ also conforms to this type of function, and similarly for the function $\exp(-k^2x^2)$. It is therefore of interest to inquire whether these analytic forms for $I(x)$ can be identified with particle size or distortion broadening or a combination of the two effects. If $h(x)$ and $g(x)$ are, however, of the form $(1+c^2x^2)^{-2}$, $I(x)$ is not of this form and cannot be readily evaluated analytically from (6). Nevertheless, for reasons to be discussed later, this form for $I(x)$ is also investigated below.

Case 1

Let $I(x) = (1+a^2x^2)^{-1}$, then from (3)

$$A(m) = P(\pi/a) \exp(-2\pi |m|/aT).$$

Hence

$$\left(\frac{d\hat{A}}{d|m|} \right)_{|m|=0} = -\frac{2\pi}{aT}, \quad \dots\dots\dots (7)$$

$$\left(\frac{d^2\hat{A}}{d|m|^2} \right)_{|m|=0} = \frac{4\pi^2}{a^2T^2}, \quad \dots\dots\dots (8)$$

$A(m)$ has a finite initial slope and therefore the function $(1+a^2x^2)^{-1}$ can represent particle size broadening. If distortion broadening were also present, from (5) and (8), \bar{e}^2 would have to be negative. Since this is physically impossible, this particular intensity distribution for $I(x)$ can only correspond with pure particle size broadening, and $A(m)$ may be replaced by $\text{const.} \times N(m)$. In such cases, Bertaut (1950) has shown that it is possible to derive useful information about the particle size and size distribution in the sample. The following results follow directly from the relationships he has established.

(i) The mean particle size \bar{M} in the $[00l]$ direction is given by

$$\bar{M} = \frac{aT}{2\pi}, \quad \dots\dots\dots (9)$$

This follows from Bertaut's analysis, which shows that

$$\overline{M} = t(0),$$

where

$$t(m) = -N(m) \left/ \left(\frac{dN}{d|m|} \right) \right|_{|m|=0} = \int_{|m|}^{\infty} (M - |m|) p(M) dM, \dots (10)$$

and $p(|m|)$ is the size distribution function.

(ii) The "apparent particle size" L as defined by Jones (1938) is given by

$$L = \int_{-\infty}^{\infty} N(m) dm / N(0) = aT/\pi, \dots (11)$$

where \overline{M} and L are in units of a_3 .

It may be noted here that, since the integral line breadth

$$\beta = \int_{-\infty}^{\infty} I(x) dx / I(0) = \pi/a,$$

it follows from (11) that

$$(\beta \cos \theta_0) / \lambda = (a_3 L)^{-1},$$

which is the familiar relationship for particle size broadening.

(iii) The mean square particle size

$$\overline{M^2} = \int_{-\infty}^{\infty} t(m) dm = a^2 T^2 / 2\pi^2 = 2(\overline{M})^2. \dots (12)$$

(iv) The mean square deviation in particle size

$$\overline{\varepsilon^2} = (aT/2\pi)^2. \dots (13)$$

(v) The size distribution function

$$p(|m|) = d^2 t / d|m|^2 = (2\pi/aT) \exp(-2\pi|m|/aT). \dots (14)$$

(vi) The fraction of particles having dimensions M in the range $|m| \leq M < \infty$ out of the total number of particles contributing to the diffraction is given by

$$\int_{|m|}^{\infty} p(M) dM = -dt/d|m| = \exp(-2\pi|m|/aT). \dots (15)$$

Case 2

Let $I(x) = \exp(-k^2 x^2)$, then

$$A(m) = P(\pi^{1/2}/k) \exp(-\pi^2 m^2 / k^2 T^2). \dots (16)$$

It follows from (4) and (16) that $\overline{M} = \infty$ if k is finite.

Hence this type of function cannot represent particle size broadening. To investigate the possibility of distortion broadening we can replace (Am) by $\text{const.} \times J(m)$.

From (5) the root mean square strain in the $[00l]$ direction is given approximately by

$$(\bar{\epsilon}^2)^{\frac{1}{2}} = (2k^2 l^2 T^2)^{-\frac{1}{2}}. \quad \dots\dots\dots (17)$$

Since $\lambda = 2d \sin \theta_0$, $d = a_3/l$, $\beta = \pi^{\frac{1}{2}}/k$, it follows by substitution in (17) that the "apparent tensile strain" η is given by

$$\eta = \beta \cot \theta_0 = 2(2\pi \bar{\epsilon}^2)^{\frac{1}{2}}. \quad \dots\dots\dots (18)$$

This is the result found by Stokes and Wilson (1944) by other methods, for distortion broadening due to a Gaussian distribution of lattice strains.

Case 3

Let $I(x) = (1 + c^2 x^2)^{-2}$, then

$$A(m) = P(\pi/2c)(1 + 2\pi |m|/cT) \exp(-2\pi |m|/cT), \quad \dots (19)$$

and

$$(d\hat{A}/d|m|)_{|m|=0} = 0.$$

Thus this function cannot represent particle size broadening. However (19) is of the correct form to represent distortion broadening. It follows that

$$(\bar{\epsilon}^2)^{\frac{1}{2}} = (clT)^{-1}, \quad \dots\dots\dots (20)$$

and

$$\eta = \pi(\bar{\epsilon}^2)^{\frac{1}{2}}. \quad \dots\dots\dots (21)$$

(b) Other Types of Particle Size Broadening

It is of interest to consider the pure diffraction contours associated with types of particle size distributions other than the somewhat unlikely case in practice given by the Cauchy contour $(1 + a^2 x^2)^{-1}$. This may be investigated by reversing the previous procedure; that is, the particle size distribution is assumed and the corresponding diffraction contour is then obtained.

By replacing $N(m)$ in (1) by the expression for $N(m)$ in terms of the size distribution function $p(|m|)$ given by (10), it follows that for a contour symmetrical about the peak value

$$I(x) = K \int_0^\infty p(M) (\pi x/T)^{-2} \sin^2(\pi M x/T) dM, \quad \dots\dots (22)$$

where, by definition,

$$\int_0^\infty p(M) dM = 1$$

and K is a constant.

This is a general expression which enables the intensity distribution for the $00l$ reflection to be determined if the size distribution function is known. In general, as Jones (1938) has pointed out, $p(M)$ is not known and is likely to vary for each material, but three possible examples are given below.

(i) Consider first the simple case in which the crystals are all of the same size and the same external shape, which is here taken to be cubic.

In this case $M = \overline{M} = L$. Hence

$$I(x) = KL^2(\pi Lx/T)^{-2} \sin^2(\pi Lx/T), \quad \dots\dots\dots (23)$$

that is, the intensity distribution is of the familiar form $D(nx)^{-2} \sin^2 nx$ where D and n are constants. The same result has been obtained by other methods by Stokes and Wilson (1942) as part of a more generalized treatment of the diffraction from polycrystalline aggregates of uniform particle size.

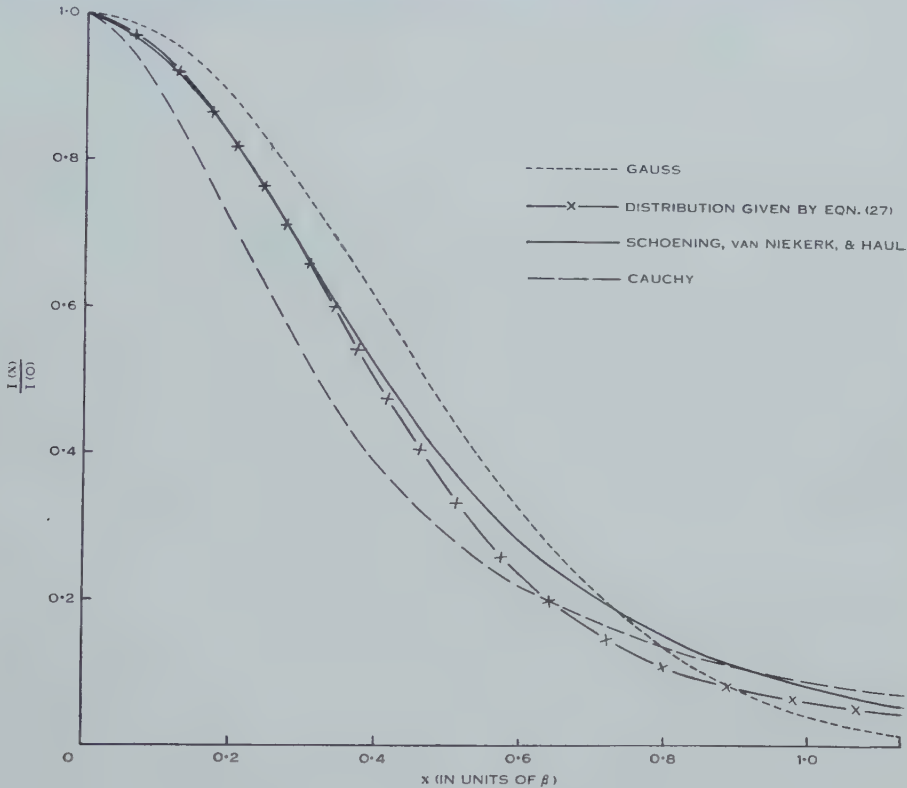


Fig. 1.—Types of distribution functions representing $I(x)$.

(ii) Let

$$p(M) = (2/\pi)^{\frac{1}{2}} \sigma^{-1} \exp(-M^2/2\sigma^2), \quad \dots\dots\dots (24)$$

where σ is the root mean square value of M .

By integration of (22) it follows that

$$I(x) = 2K(2\pi x/T)^{-2} [1 - \exp\{-\frac{1}{2}\sigma^2(2\pi x/T)^2\}]. \quad \dots\dots (25)$$

(iii) One objection to (24) is that $p(M)$ is a maximum for $M=0$. To obtain a skew distribution about a non-zero value of M let

$$p(M) = (2/\pi)^{\frac{1}{2}} S^{-3} M^2 \exp(-M^2/2S^2), \quad \dots\dots\dots (26)$$

where S is a constant. Then

$$I(x) = 2K(2\pi x/T)^{-2} [1 - (1 - 4\pi^2 S^2 x^2/T^2) \exp\{-\frac{1}{2}S^2(2\pi x/T)^2\}]. \quad \dots\dots\dots (27)$$

TABLE 1
PURE DIFFRACTION CONTOURS ASSOCIATED WITH DIFFERENT CAUSES OF LINE BROADENING

Pure Diffraction Contour $I(x)$	Type of Broadening	Size Distribution Function $p(M)$	" Apparent Particle Size " L	Mean Particle Size* \bar{M}
$(1+a^2x^2)^{-1}$	Particle size	$\frac{2\pi}{aT} \exp\left(-\frac{2\pi M}{aT}\right)$	$\frac{aT}{\pi}$	$\frac{aT}{2\pi}$
$(nx)^{-2} \sin^2 nx$	"	Dirac delta function : $\delta\left(M - \frac{nT}{\pi}\right)$	$\frac{nT}{\pi}$	$\frac{nT}{\pi}$
$\frac{T^2}{2\pi^2\sigma^2x^2} \left(1 - \exp\left(-\frac{2\pi^2\sigma^2x^2}{T^2}\right)\right)$	"	$\left(\frac{2}{\pi}\right)^{\frac{1}{2}} \frac{1}{\sigma} \cdot \exp\left(-\frac{M^2}{2\sigma^2}\right)$	$\left(\frac{\pi}{2}\right)^{\frac{1}{2}} \sigma$	$\left(\frac{2}{\pi}\right)^{\frac{1}{2}} \sigma$
$\frac{T^2}{6\pi^2S^2x^2} \left\{1 - (1 - 4\pi^2S^2x^2) \exp\left(-\frac{2\pi^2S^2x^2}{T^2}\right)\right\}$	"	$\left(\frac{2}{\pi}\right)^{\frac{1}{2}} \frac{1}{S^3} \cdot M^2 \exp\left(-\frac{M^2}{2S^2}\right)$	$3\left(\frac{\pi}{8}\right)^{\frac{1}{2}} S$	$\left(\frac{8}{\pi}\right)^{\frac{1}{2}} S$
Distortion $\exp(-k^2x^2)$ $(1+c^2x^2)^{-2}$	"	" Apparent Tensile Strain " η		
		$2(2\pi c^2)^{\frac{1}{2}}$ $\pi(c^2)^{\frac{1}{2}}$		

* Particle sizes are expressed in units of axial length a_0 .

The results of the preceding analysis are summarized in Table 1 and in Figures 1 and 2. In these figures, the constants for each function have been adjusted for convenience so as to make the areas under the curves all equal.

(c) *Combined Particle Size and Distortion Broadening*

In the examples considered so far, either particle size or distortion broadening has alone been operative. Both effects may however be present in the same sample and Warren and Averbach (1952*b*) have shown how the experimental

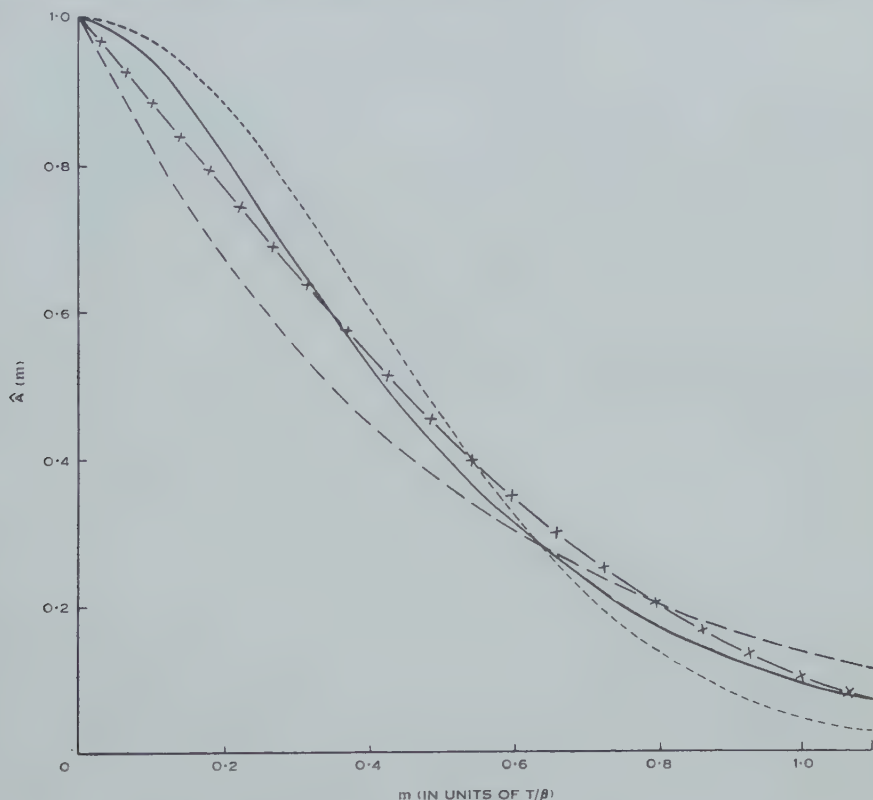


Fig. 2.—Transforms of functions in Figure 1.

data may, in favourable cases, be analysed numerically to determine the relative influence of the two factors. It is, however, of interest to reverse this procedure and examine analytically how particular models of size and distortion combinations may be expected to modify the shape of the function $A(m, l)$ for any given reflection.

The problem is of particular importance in connexion with the effect of cold-work on polycrystalline metals. On the assumption that both particle size and distortion broadening are contributory factors, a number of models may be assumed. Two extreme cases are as follows:

(1) First, as a result of plastic deformation, each grain in the aggregate may become dissociated into a number of units. Some of these units are strain free but of such dimensions that particle size broadening occurs; in other units

the size is still sufficiently large for no appreciable size broadening, but distortion broadening occurs due to heterogeneous lattice strains. Under these circumstances the intensity distribution in a reflection is the sum of the contributions due to the particle size and distortion factors treated independently. This would correspond with the assumption made by Hall (1949) in taking the total line breadth β as the simple sum of the breadths due to each factor separately.

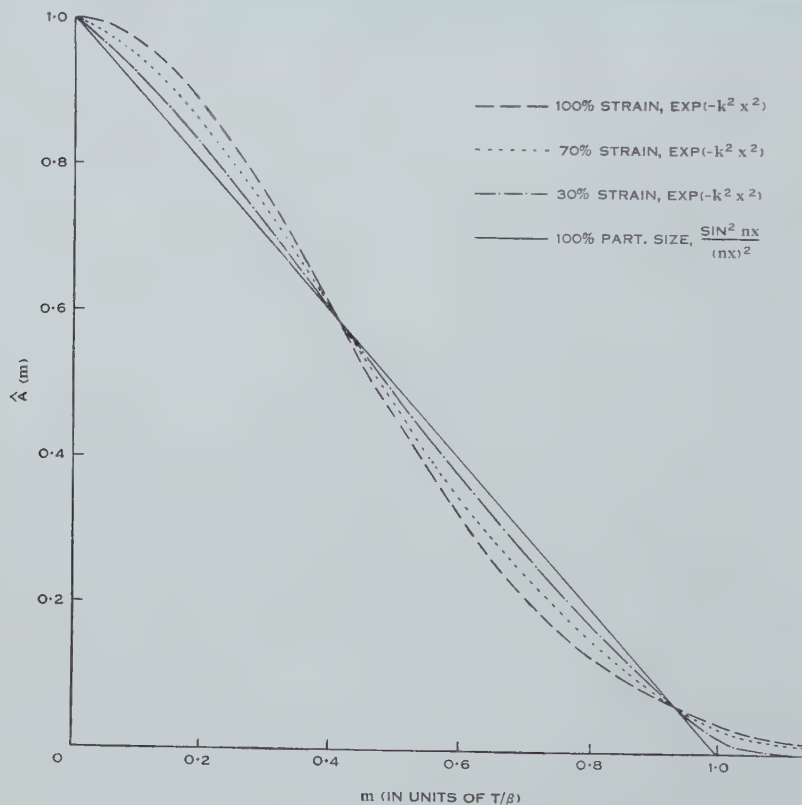


Fig. 3.—Transforms for combined particle size and distortion broadening given by equation (29).

(2) Secondly and more probably, after plastic deformation each grain may become dissociated into a number of small, heterogeneously distorted units. The intensity distribution for any reflection is then the resultant coherent scattering in a given direction of the incident beam produced by these domains, and is given by equation (1) for crystals of the cubic class.

As an illustration of the method of treatment, a particular example corresponding to each of these cases is considered below.

Case 1

On the first hypothesis, suppose for simplicity that the crystallites conform to one or the other of only two possible types:

(i) A given proportion of the total number contains crystallites sufficiently large for no size broadening to occur and in which the strain distribution function

is assumed to be Gaussian. The intensity distribution in the reflected beam due to these crystallites is thus of the form $\exp(-k^2x^2)$.

(ii) The remainder are strain free and of a small uniform size. For these, the diffraction contour is represented by the function $(nx)^{-2} \sin^2 nx$.

The resultant pure diffraction contour is then given by

$$I(x) = C \exp(-k^2x^2) + D(nx)^{-2} \sin^2 nx. \quad \dots\dots (28)$$

Hence

$$A(m) = K_1(\pi^{1/2}/k) \exp(-\pi^2 m^2/k^2 T^2) + K_2(\pi/n)(1 - \pi |m|/nT)^* \dots (29)$$

where K_1 and K_2 are parameters whose relative magnitudes determine the contributions due to distortion and particle size effects respectively. For 100 per cent. strain broadening $K_1=1$, $K_2=0$ and for 100 per cent. size broadening $K_1=0$, $K_2=1$.

Figure 3 shows the form of $\hat{A}(m)$ for values of $(K_1, K_2) = (1, 0)$, $(0.7, 0.3)$, $(0.3, 0.7)$, and $(0, 1)$.

Case 2

In this example, it is assumed that the sample consists of small distorted crystals of uniform size and with a Gaussian distribution of lattice strains. Hence

$$A(m) = N(m)J(m) = P(1 - \pi |m|/nT) \exp(-\pi^2 m^2/k^2 T^2)^* \dots (30)$$

where P is a constant and n and k have the same significance as in the previous example. By assigning various values to n and k varying relative contributions of particle size and distortion broadening respectively can be represented. For pure size broadening, $k=\infty$ and for pure strain broadening $n=\infty$. The true diffraction line breadth is given by

$$\beta^{-1} = \pi^{-1/2} kT \operatorname{erf}(n/k) - (k^2 T/\pi n) \{1 - \exp(-n^2/k^2)\}. \dots (31)$$

In Figure 4 $\hat{A}(m)$ is plotted for values of n/k of 0, 2, 5, ∞ . The curves all represent the same value for β .

(d) Correction Curves for Particle Size Broadening

To obtain the true integral breadth β from the measured values B and b it is convenient and customary to obtain, if possible, correction curves in which β/B may be plotted as functions of b/B . The principal correction formulae which have been suggested are summarized in Table 2. As pointed out in Section I, these formulae are based upon the subjective choice of various types of analytic functions to represent the experimental X-ray line profiles, and in general are without reference to the particular cause of broadening involved. The preceding analysis, however, indicates how given types of analytic functions

* The particle size factor in (29) and (30) is zero outside the range $0 \leq |m| \leq nT/\pi$.

may be assigned to particular forms of particle size distributions. It is therefore of interest to investigate the dependence of the appropriate correction curve upon the particular type of particle size broadening considered.

The appropriate relationship between B , b , and β can, however, only be obtained if two of the three functions in equation (6) above are known. Assuming

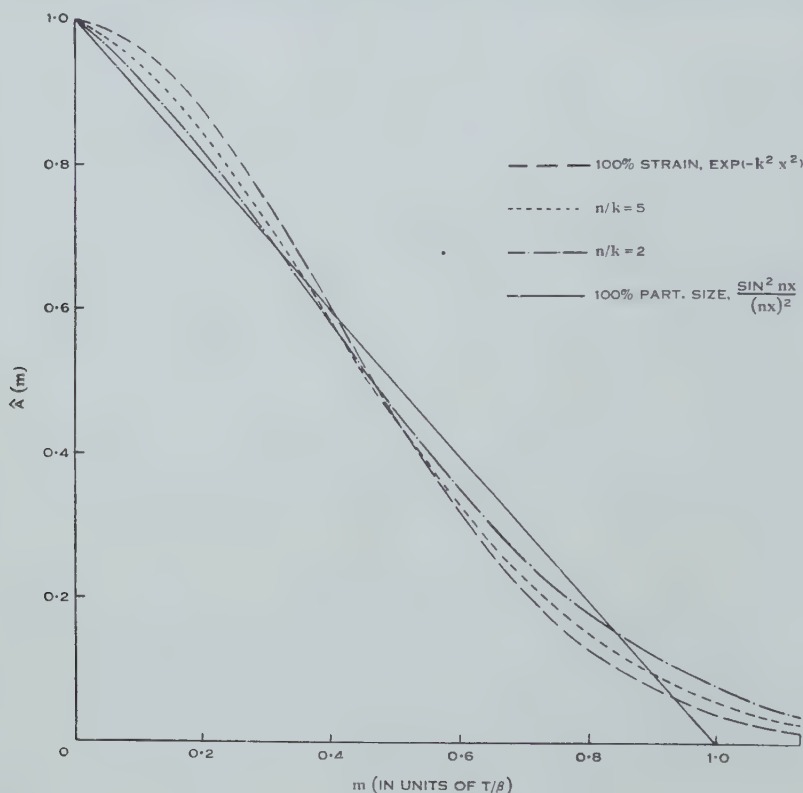


Fig. 4.—Transforms for combined particle size and distortion broadening given by equation (30).

that $I(x)$ and $g(x)$ are known, the integral breadth B may be obtained by either of the following two methods.

(i) From (6) it follows that if transforms of $h(x)$ and $g(x)$ are given by

$$H(m) = \int_{-\infty}^{\infty} h(x) \exp(2\pi i m x) dx,$$

$$G(m) = \int_{-\infty}^{\infty} g(x) \exp(2\pi i m x) dx$$

respectively, then

$$\begin{aligned} H(m) &= \bar{A}(Tm)G(m), \\ H(0) &= G(0). \end{aligned}$$

TABLE 2
RELATIONSHIP BETWEEN B , b , AND β FOR VARIOUS LINE PROFILES

Authors	Correction Formula	Type of Function Representing Experimental Profiles $g(x)$ and $h(x)$ for Correction Formula to be Valid
(I) Scherrer (1920)	$\beta = B - b$	$(1 + a^2 x^2)^{-1}$
(II) Warren and Biscoe (1938) ..	$\beta = (B^2 - b^2)^{\frac{1}{2}}$	$\exp(-k^2 x^2)$
(III) Jones (1938)	Correction curves relating β/B to b/B	Shape of $g(x)$ determined experimentally, $I(x)$ assumed to be of type $(1 + a^2 x^2)^{-1}$ or $\exp(-k^2 x^2)$
(IV) Taylor (1941)	$\beta = [(B - b)(B^2 - b^2)^{\frac{1}{2}}]^{\frac{1}{2}}$	No assumptions as to $g(x)$ and $h(x)$ but β taken arbitrarily as geometric mean of (I) and (II)
(V) Alexander and Klug (1950)	Correction curves relating β/B to b/B	$I(x)$ assumed to be $(1 + a^2 x^2)^{-1}$ and X-ray source contour of form $\exp(-k^2 x^2)$
(VI) Schoening, van Niekerk, and Haul (1952)	Correction curves relating β/B to b/B	$(1 + c^2 x^2)^{-2}$

TABLE 3
CORRECTION FORMULAE FOR DIFFERENT TYPES OF PARTICLE SIZE BROADENING AND DIFFERENT INSTRUMENTAL CONTOURS $g(x)$

$p(M)$	β	Relationship between B , b , and β	
		$g(x) = \exp(-k^2x^2)$, $b = \pi^{1/2}/k$	$g(x) = (1 + c^2x^2)^{-2}$, $b = \pi/2c$
$\frac{2\pi}{aT} \exp\left(-\frac{2\pi}{aT}M\right)$	$\frac{\pi}{a}$	$\frac{b}{B} = \exp \frac{\beta^2}{\pi b^2} \operatorname{erfc} \frac{\beta}{\pi^{1/2}b}$	$\frac{\beta}{B} = \frac{1}{2} - \frac{2b}{B} + \frac{1}{2} \left(\frac{8b}{1+B} \right)^{1/2}$
$\delta \left(M - \frac{nT'}{\pi} \right)$	$\frac{\pi}{n}$	$\frac{b}{B} = \operatorname{erf} \frac{\pi^{1/2}b}{\beta} - \frac{\beta}{\pi b} \left[1 - \exp \left(-\frac{\pi b^2}{\beta^2} \right) \right]$	$\frac{b}{B} = \frac{\beta}{8b} \left[\left(3 + \frac{4b}{\beta} \right) \exp \left(-\frac{4b}{\beta} \right) - 3 + \frac{8b}{\beta} \right]$
$\left(\frac{2}{\pi} \right)^{1/2} \cdot \frac{1}{\sigma} \cdot \exp \left(-\frac{M^2}{2\sigma^2} \right)$	$\left(\frac{2}{\pi} \right)^{1/2} \cdot \frac{T}{\sigma}$	$\frac{\beta}{B} = 1 - \frac{b^2}{B^2}$	$\frac{\beta}{B} = -\frac{3\pi\beta^2}{16b^2} + \frac{3\beta}{2b} + \left(\frac{3\pi\beta^2}{16b^2} - 2 \right) \exp \frac{16b^2}{\pi\beta^2} \operatorname{erfc} \frac{4b}{\pi^{1/2}\beta}$
$\left(\frac{2}{\pi} \right)^{1/2} \cdot \frac{M^2}{S^3} \exp \left(-\frac{M^2}{2S^2} \right)$	$\left(\frac{8}{\pi} \right)^{1/2} \cdot \frac{T}{3S}$	$\frac{b}{B} = \frac{1}{8b} \left[\frac{9\beta^2 + 32b^2}{(9\beta^2 + 16b^2)^{1/2}} - 3\beta \right]$	$\frac{\beta}{B} = -\frac{9\pi\beta^2}{64b^3} + \frac{3\beta}{4b} + \frac{32b}{9\pi\beta} + \left(\frac{9\pi\beta^2}{64b^2} - \frac{256b^2}{27\pi\beta^2} \right) \exp \frac{64b^2}{9\pi\beta^2} \operatorname{erfc} \frac{8b}{3\pi^{1/2}\beta}$

The observed intensity distribution $h(x)$ in the presence of size broadening is then found from

$$h(x) = \int_{-\infty}^{\infty} H(m) \exp(-2\pi i m x) dm.$$

Finally

$$B = \int_{-\infty}^{\infty} h(x) dx / h(0) = H(0) / h(0),$$

or

$$B = G(0) / \int_{-\infty}^{\infty} H(m) dm. \quad \dots\dots\dots (32)$$

(ii) The above method is appropriate if the pure diffraction contour $I(x)$ is known directly. If, however, the size distribution function $p(|m|)$ is adopted as a starting-point, it is more convenient to derive an expression for the breadth B in terms of this distribution function. The analysis is too lengthy to set out here, but it can be shown that

$$B = \overline{MG}(0) / \int_{-\infty}^{\infty} t(Tm)G(m)dm, \quad \dots\dots\dots (33)$$

where the parameters have the significance defined previously.

As pointed out already, B will depend upon the functions chosen to represent $I(x)$ and $g(x)$. In practice, it is generally found that the instrumental contour $g(x)$ may be represented quite closely, either by $\exp(-k^2x^2)$ (Taylor and Sinclair 1945; Shull 1946; Alexander 1950), or by a function of the type $(1+c^2x^2)^{-2}$ (Jones 1938; Schoening, van Niekerk, and Haul 1952).

For both of these forms for $g(x)$, the four types of particle size broadening considered previously have been analysed by the methods outlined above and the results are summarized in Table 3. In Figure 5, the correction curves corresponding to the size distribution function $p(M)$ given by (26) have been plotted in the usual way for the two different forms for $g(x)$. For comparison, the curves corresponding to formulae (I), (II), and (VI) in Table 2 are also included in Figure 5.

III. DISCUSSION

A number of interesting points emerge from the analysis in Section II.

(1) First, the functions $(1+a^2x^2)^{-1}$ and $\exp(-k^2x^2)$, which have previously been used somewhat indiscriminately to represent the pure diffraction contour, can in fact only be identified with line broadening due to *particular* types of particle size and lattice distortion effects respectively. Hence, in any attempt to distinguish between particle size or strain broadening from a particular material, the use of the one or the other of these functions (together with the appropriate relationship between B , b , and β) involves an intrinsic initial assumption about the cause of the broadening, when the object of the investigation is to discover the cause. Such an assumption must inevitably weight the experimental results, partially at least, in favour of one or the other of the two effects.

In this connexion it is therefore perhaps significant that in most previous work on the cause of line broadening from cold-worked metals, those investigators who have used the Warren relationship between B , b , and β have concluded that lattice distortion was the predominant factor, whilst those who have employed the Scherrer correction found that particle size was the main cause. The best procedure in such work therefore is to make no assumptions at all about the shape of the experimental line profiles.

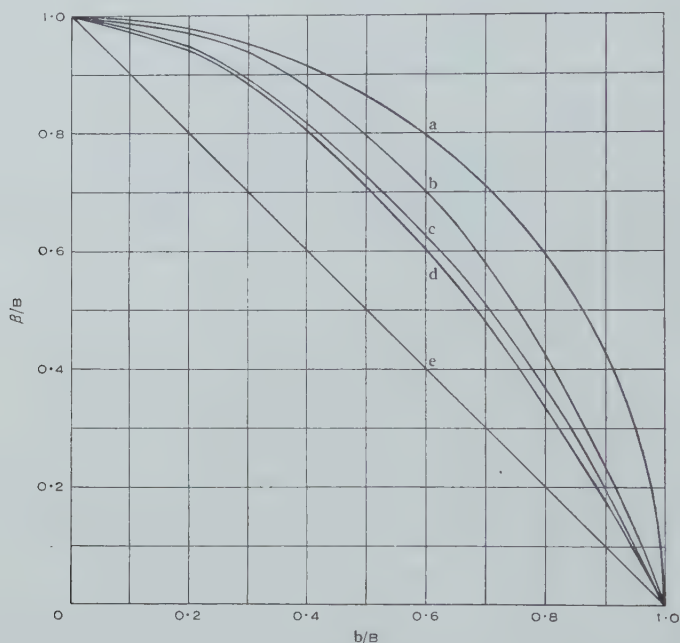


Fig. 5.—Correction curves associated with particle size broadening for various types of size distribution functions and different instrumental contours. *a*, Warren curve; *b*, $p(M)$ given by equation (26), $g(x) = \exp(-k^2x^2)$; *c*, $p(M)$ given by equation (26), $g(x) = (1 + c^2x^2)^{-2}$; *d*, Schoening, van Niekerk, and Haul curve; *e*, Scherrer curve.

(2) Secondly, as would be expected, the pure diffraction contour depends not only upon the cause of the broadening but also upon the nature of the size and strain distribution functions. In cases where the size effect is negligible, Eastabrook and Wilson (1952) have suggested that the Cauchy line profile is possibly a better approximation to the pure diffraction contour encountered in practice than the Gaussian form. At first sight this is in direct contradiction to the results obtained here. It should be noted, however, that different distortion models are considered in the two cases. In the present paper it has been shown that a Gaussian strain distribution corresponds to a Gaussian intensity distribution for the pure diffraction contour. On the other hand, Eastabrook and Wilson have considered the case of crystals large enough for negligible particle size broadening and in which each crystal is *assumed to contain several regions of compression and extension*. In this case, they have shown that for small m

(corresponding to the behaviour of $I(x)$ for large x), $J(m)$ is approximately proportional to $\exp(-2\pi^2 l^2 \bar{e}^2 m^2)$, whereas for large m , $J(m)$ tends more nearly to be proportional to $\exp(-\text{const.} |m|)$. This implies that, for this particular type of distortion broadening, the pure diffraction contour would approximate to a Cauchy form at and near to its peak value, with a gradual transition with progressive increase in x to some other profile (e.g. Gaussian) that is consistent with distortion broadening for small m .

The question as to whether the contour is a closer resemblance to a Cauchy or to a Gaussian profile will depend upon the type of distortion that is operative. For cases of distortion broadening encountered in practice, it seems unlikely that any simple analytic function will truly represent the pure diffraction contour, and for the reasons discussed in (1) above and (4) below, it is dangerous to attempt to assign given functions as "close fits" to the experimental observations when dealing with distortion broadening. For pure particle size broadening on the other hand, such a procedure is often permissible, since determination of the "apparent particle size" from line breadth measurements can only be regarded as approximate in view of a number of other uncertainties.

(3) Thirdly, for pure particle size broadening, Alexander (1950), in discussing the factors affecting determination of crystallite size with a Geiger-counter X-ray spectrometer, concluded that a typical crystal size distribution leads to a pure diffraction contour which bears a closer resemblance to a Cauchy profile than a Gaussian form. The preceding analysis would tend to support this conclusion. At the same time the exact form of the contour depends upon the particular type of size distribution law appropriate to each material.

A point of interest here is that Schoening, van Niekerk, and Haul (1952) have reported recently that the line profiles obtained from small crystals with a Geiger-counter X-ray spectrometer could be represented very closely by functions of the type $(1+c^2x^2)^{-2}$. In this case, although $I(x)$ cannot be conveniently evaluated analytically, by employing the methods described previously it can readily be shown that these functions for $h(x)$ and $g(x)$ cannot truly represent particle size broadening. Assuming that in the samples examined by these authors no distortion broadening occurred, the inconsistency between their experimental observations and the present theoretical treatment presents a problem. It may perhaps be that some other function or functions compatible with particle size broadening could also represent closely their experimental observations. For example, if

$$g(x)/g(0) = (1 + q_1 c_1^2 x^2)(1 + c_1^2 x^2)^{-2}$$

and

$$h(x)/h(0) = (1 + q_2 c_2^2 x^2)(1 + c_2^2 x^2)^{-2}$$

(which approximate to the previous functions if q_1 and q_2 are small), it can be shown that these profiles are compatible with size broadening, giving a non-negative value for the size distribution function $p(|m|)$ provided that

$$0 \leq \frac{1-q_2}{1+q_2} \cdot \frac{1}{c_2} \leq \frac{1-q_1}{1+q_1} \cdot \frac{1}{c_1} + \frac{1}{2} \left(\frac{1}{c_2} - \frac{1}{c_1} \right)^2 \left/ \left(\frac{1}{c_2} - \frac{2q_1}{1+q_1} \cdot \frac{1}{c_1} \right) \right.$$

Examination of Figure 5, however, shows that except in cases where a high accuracy is warranted in the determination of particle size, the differences introduced by using one or other of the correction curves (b), (c), or (d) will not be serious. It is also interesting to note the close agreement between curves (c) and (d) which are both based upon the same function $(1+c^2x^2)^{-2}$ for $g(x)$. Curve (d), given by Schoening, van Niekerk, and Haul, cannot strictly be applicable to particle size broadening, yet is nevertheless a close fit to *their* experimental conditions and observations. On the other hand curve (c) corresponds to a particular size distribution which may approximate to typical distributions encountered in practice. It appears therefore that for general application to particle size broadening, curve (c) is likely to be a reasonable compromise.

(4) Finally, as pointed out by Eastabrook and Wilson (1952), the original methods developed by Warren and Averbach (1950) for distinguishing between particle size and distortion effects, based upon the shape of the pure diffraction contour rather than its breadth, have certain limitations in practice, particularly if the analysis is restricted to single orders of reflection from given sets of planes in the crystal. It is true, for example, that there is a considerable difference in shape between the Fourier transforms (Fig. 2) corresponding to the contour functions $(1+a^2x^2)^{-1}$ and $\exp(-k^2x^2)$. Comparison of Figures 1 and 2, however, shows that the pure diffraction contours in Figure 1, which produce appreciable changes in Figure 2, only differ appreciably in shape in regions close to their "tails" where percentage errors in measurement are greatest. In consequence, if Fourier methods are used to determine the pure diffraction contour $I(x)$ or its Fourier coefficients, from observations on the shapes of the experimental $h(x)$ and $g(x)$ contours, the shape of the $A(m)$ curve will be critically dependent upon the accuracy of the experimental observations of line intensities at values approaching the general background value due to incoherent scattering. Furthermore, Figures 3 and 4 illustrate how the shape of the curve changes when particle size and distortion effects are both operative. Consequently, although the more recent methods of Warren and Averbach (1952*b*), based on analysis of several orders of reflection from any given set of planes, reduce the chance of ambiguity of interpretation (due to experimental error) of the experimental results, the use of such criteria *alone*, for differentiating between particle size and distortion broadening, would appear to be rather dangerous unless very accurate experimental techniques are employed.

IV. ACKNOWLEDGMENT

The authors desire to thank the Chief Scientist, Department of Supply, Australia, for permission to publish this paper.

V. REFERENCES

- ALEXANDER, L. (1950).—*J. Appl. Phys.* **21**: 126.
- ALEXANDER, L., and KLUG, H. P. (1950).—*J. Appl. Phys.* **21**: 137.
- AULD, J. H., and GARROD, R. I. (1952).—*Nature* **169**: 579.
- BERTAUT, E. F. (1950).—*Acta Cryst. Camb.* **3**: 14.
- BERTAUT, E. F. (1952).—*Acta Cryst. Camb.* **5**: 117.

- BRINDLEY, G. W. (1940).—*Proc. Phys. Soc. Lond.* **52**: 117.
- EASTABROOK, J. N., and WILSON, A. J. C. (1952).—*Proc. Phys. Soc. Lond.* B **65**: 67.
- HALL, W. H. (1949).—*Proc. Phys. Soc. Lond.* A **62**: 741.
- JONES, F. W. (1938).—*Proc. Roy. Soc. A* **166**: 16.
- McKEEHAN, M., and WARREN, B. E. (1953).—*J. Appl. Phys.* **24**: 52.
- MEGAW, HELEN D., and STOKES, A. R. (1945).—*J. Inst. Met.* **71**: 279.
- PATERSON, M. S. (1950).—*Proc. Phys. Soc. Lond.* A **63**: 477.
- PATERSON, M. S. (1952).—*J. Appl. Phys.* **23**: 499.
- SCHERRER, P. (1920).—In Zsigmondy, R., "Kolloidchemie." 3rd Ed. p. 387. (Spamer: Leipzig.)
- SCHOENING, F. R. L., VAN NIEKERK, J. N., and HAUL, R. A. W. (1952).—*Proc. Phys. Soc. Lond.* B **65**: 528.
- SHULL, C. G. (1946).—*Phys. Rev.* **70**: 679.
- SMITH, C. S., and STICKLEY, E. E. (1943).—*Phys. Rev.* **64**: 191.
- SMITH, R. A. (1953).—*J. Iron St. Inst.* **173**: 147.
- STOKES, A. R. (1948).—*Proc. Phys. Soc. Lond.* **61**: 382.
- STOKES, A. R., PASCOE, K. J., and LIPSON, H. (1943).—*Nature* **151**: 137.
- STOKES, A. R., and WILSON, A. J. C. (1942).—*Proc. Camb. Phil. Soc.* **38**: 313.
- STOKES, A. R., and WILSON, A. J. C. (1944).—*Proc. Phys. Soc. Lond.* **56**: 174.
- TAYLOR, A. (1941).—*Phil. Mag.* **31**: 339.
- TAYLOR, A., and SINCLAIR, H. (1945).—*Proc. Phys. Soc. Lond.* **57**: 126.
- WARREN, B. E., and AVERBACH, B. L. (1950).—*J. Appl. Phys.* **21**: 595.
- WARREN, B. E., and AVERBACH, B. L. (1952a).—"Imperfections in Nearly Perfect Crystals." p. 152. (John Wiley & Sons: New York.)
- WARREN, B. E., and AVERBACH, B. L. (1952b).—*J. Appl. Phys.* **23**: 497.
- WARREN, B. E., and BISCOE, J. (1938).—*J. Amer. Ceram. Soc.* **21**: 49.
- WILLIAMSON, G. K., and HALL, W. H. (1953).—*Acta Metallurgica* **1**: 22.
- WOOD, W. A., and RACHINGER, W. A. (1949).—*J. Inst. Met.* **75**: 571.

GALACTIC RADIATION AT RADIO FREQUENCIES

VII. DISCRETE SOURCES WITH LARGE ANGULAR WIDTHS

By J. G. BOLTON,* K. C. WESTFOLD,† G. J. STANLEY,* and O. B. SLEE*

[*Manuscript received September 25, 1953*]

Summary

Observations with three forms of equipment have revealed the existence of a number of sources of angular width more than 1° .

A rough analysis of the brightness distribution of one source shows that it is elongated along a parallel of galactic latitude. This source appears to be typical of a class that is generally distributed around the galactic equator, and may represent fine structure in the distribution of radiation from the Galaxy.

Of the others, one appears to be associated with the abnormal galaxy NGC 5128, and another has been identified with a network of gaseous filaments in our own Galaxy.

I. INTRODUCTION

Up to the present time, surveys of the brightness distribution of galactic noise have been made using either a single aerial of low resolving power or an interferometer consisting of two aerials separated by many wavelengths. The observations have been mainly on metre wavelengths, for which the brightness is sufficiently high to permit measurements over a large part of the celestial sphere.

With such wavelengths, aerials of moderate physical dimensions have beam widths of the order of 10° between half-power points. The observed brightness distributions contain only such major features as the concentration towards the galactic plane with its principal maximum near the galactic centre, and subsidiary maxima in Cygnus and Taurus. The actual brightness distribution can, to some extent, be extracted mathematically from the observed distribution, given the aerial sensitivity pattern, but the manual procedure is most laborious and open to errors. Higher resolution due to the sharp-edged shadow of the Earth can be gained if observations are made with an aerial on a high cliff directed at the horizon.

With the interference techniques the aerial beam is split into a number of closely spaced fringes. As the source passes through the fringe system a sinusoidal pattern is recorded, whose features are determined by the angular distribution of brightness across the source. Most interferometers used so far have aerial spacings of about 50 wavelengths, giving fringe separations of about

* Division of Radiophysics, C.S.I.R.O., University Grounds, Sydney.

† University of Sydney.

1°. The sources discovered with such systems have angular widths of a few minutes of arc. No appreciable interference pattern is obtained from a source whose angular diameter is of the order of the fringe separation.

No systematic search has yet been made for objects of intermediate size, although a number are known to exist. Mills (1952*a*) found evidence for three such sources from a survey made with two interferometers with different aerial spacings. An extended object in Cygnus has been studied by Brown and Hazard (1951*a*) at 158 Mc/s and by Piddington and Minnett (1952) at 1200 Mc/s.

The present paper describes a survey of the sky for sources whose angular widths are 1° or more. The principal equipment used was a two-aerial interferometer whose fringe spacing could be varied from 3 to 14°. The first observations with the smallest fringe spacings indicated the existence of a few sources which had not been detected with interferometers of fringe spacings of about 1°. If the sources had been sufficiently separated, it would have been possible to gain some idea of the brightness distribution across these sources by observing the amplitudes of the interference patterns at a number of fringe spacings. However, as the fringe spacing was progressively increased, the simple patterns of the smaller spacings gave place to extremely complex patterns, apparently the result of having more than one such source in the fringe pattern at a time. Such complex effects were observed whenever the plane of the Galaxy crossed the fringe system, and may perhaps be attributed to fine structure of the background. Only in relatively few cases has it been possible to suggest a specific interpretation of the observations.

II. EQUIPMENT USED IN THE INVESTIGATION OF THE SOURCES

Three types of equipment have been used to study the sources. These are a 72-ft diameter fixed reflector, a sea interferometer with automatic control of the receiver gain, and an azimuth interferometer. The last two have been described in a previous paper (Bolton and Slee 1953).

(a) *The 72-ft Reflector*

This reflector consists of $\frac{1}{2}$ in. metal strips spaced 1 ft apart on the surface of a paraboloid of revolution, of focal length 40 ft and aperture 72 ft, with its axis vertical. The feed is a rod dipole and parasitic reflector. The aerial beam width is 6° at a frequency of 150 Mc/s and the beam can be directed by tilting the mast supporting the feed. This aerial has about the same ratio of focal length to aperture as the 220-ft reflector at Jodrell Bank, for which Brown and Hazard (1951*b*) found very little distortion in the beam for tilt angles of up to 15°. By tilting up to 15°, a survey was made of the strip of the celestial sphere between declinations -20 and -50° and Right Ascensions 14 and 22 hr; this includes the galactic equator between longitudes 300 and 335°. The observations were made over a period of only a few months and electrical interference during the day-time prevented satisfactory observations of the other section of the galactic equator. The results of this survey, expressed in the form of contours of equal aerial temperature, are shown in Figure 1.

(b) *The Sea Interferometer with Automatic Control of the Receiver Gain*

The aerial of this equipment consists of a 6 by 2 array of Yagis, on an azimuth mounting erected on a 240-ft cliff overlooking the sea. It operates on a frequency of 110 Mc/s and the beam width in azimuth is about 10° . The beam width in altitude is also about 10° but the effect of the Earth's shadow at the horizon greatly improves the resolving power of the aerial, so that a faint source rising above the horizon produces a more marked change in the recorded output of the receiver than it would when crossing the free-space pattern of the aerial. The aerial with its image in the sea forms an interferometer and, if the angular width of a source is smaller than the fringe separation (1°), an interference pattern is observed.

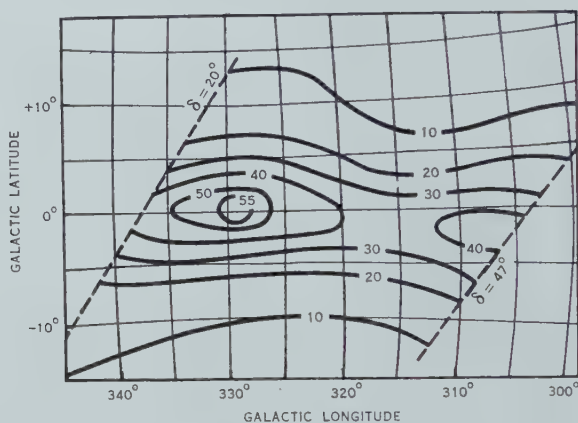


Fig. 1.—Contours of aerial temperature of the region about the galactic centre as seen on the 72-ft reflector at 160 Mc/s (aerial beam width 6° between half-power points). The units are not accurately known owing to uncertainty in the reflection coefficient of the aerial, but the contour interval is approximately 20°K .

Normally, with the sea interferometer it is difficult to distinguish the effects of faint sources against the large, but slowly varying, changes in the output of the receiver as the regions near the galactic plane cross the aerial beam. However, with a receiver modification described by Bolton and Slee (1953) this disadvantage has been largely overcome. In this modification, which is a form of automatic gain control, the output of the receiver is fed into an integrator, the output of which is in turn used to control the gain of the input stages of the receiver. The result of this arrangement is to suppress almost entirely the slowly varying components in the receiver output. Sources of angular width less than the fringe spacing appear as an interference pattern, and those of angular width somewhat greater than the fringe spacing as a small "hump", on the record. Such sources were roughly located by the sidereal time at which they first appeared, the estimated time at which they were fully risen, and a knowledge of the aerial diagram in azimuth.

(c) The Azimuth Interferometer

This equipment and some of its uses have been described in a previous paper (Bolton and Slee 1953). It consists of two aerials, whose spacing can be varied, on the top of a cliff overlooking the sea. The frequency used was 100 Mc/s. The aerial beam is split into a double system of interference fringes, due to the spacing of the two aerials and their images in the sea. The present observations were made with aerial spacings along the cliff top of between 4 and 20 wavelengths. The spacing between the aerials and their images was 50 wavelengths. Some of the observations were made with a stationary fringe system, and in others the fringe system was swung backwards and forwards through one fringe width per minute. The fringe swinging was found particularly useful with small aerial spacings where the movement of a source through the fringes, due to the Earth's rotation, is comparatively slow.

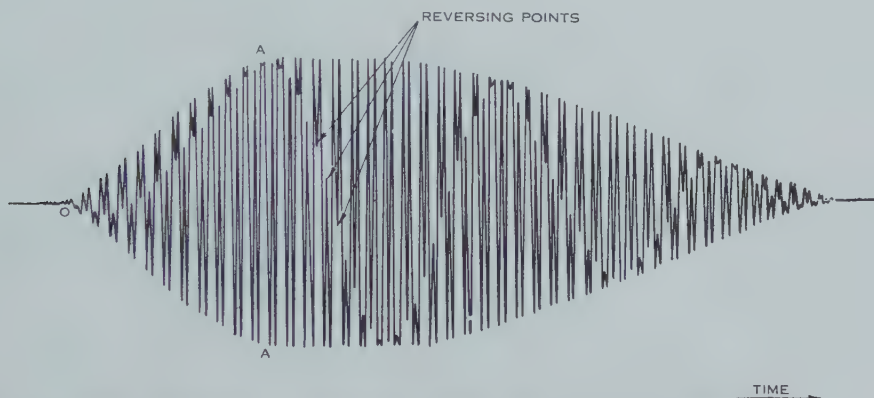


Fig. 2.—Idealized pattern due to an extended source rising above the horizon and through the fringe system of the azimuth interferometer. The fringe system is swung backwards and forwards through 1 fringe-width/min. The points at which the direction of the fringe swing is reversed can be clearly seen. From *O* to *A* the amplitude of the pattern increases as the source rises above the horizon, and then slowly decreases as the source passes out of the aerial beam. Due to phase switching of the aerials the pattern is distributed about a central zero line. Note also the change in the period of the pattern drawn by the reversing points; this is due to the motion of the instantaneous centre of the source during and after rising.

For sources whose angular widths in altitude exceed the separation of the sea-interference fringes, the effect of the azimuth fringe system only is observed. An idealized record with fringe swinging, due to such a source, is shown in Figure 2. The instantaneous amplitude of the pattern between the points marked *O*, when the source starts to appear above the horizon, and *A*, when it is fully risen, represents the total flux density from that part of the source above the horizon. The subsequent decrease in the amplitude of the pattern is due to the source passing out of the aerial beam. Apart from a small effect as the rising edge of the source enters the aerial beam, the envelope of the pattern from *O* to *A* gives the flux density per unit altitude from the source, along a line perpendicular to the horizon. The flux density per unit azimuth, along a line parallel to the horizon, can be deduced from the variation of the pattern amplitude with aerial spacing.

The points at which the fringe swing is reversed at each end of the cycle are clearly shown on Figure 2. The locus of these points represents one of the interference patterns that would be obtained with the stationary fringe system. The period of this pattern is due to the azimuth motion of the effective centre of the source and, ideally, it can be used to determine its declination. However, while the source is rising, the period of the locus is often different from that when it is fully above the horizon. This effect is due to the motion of the instantaneous centre of the visible part of the source. For a source rising

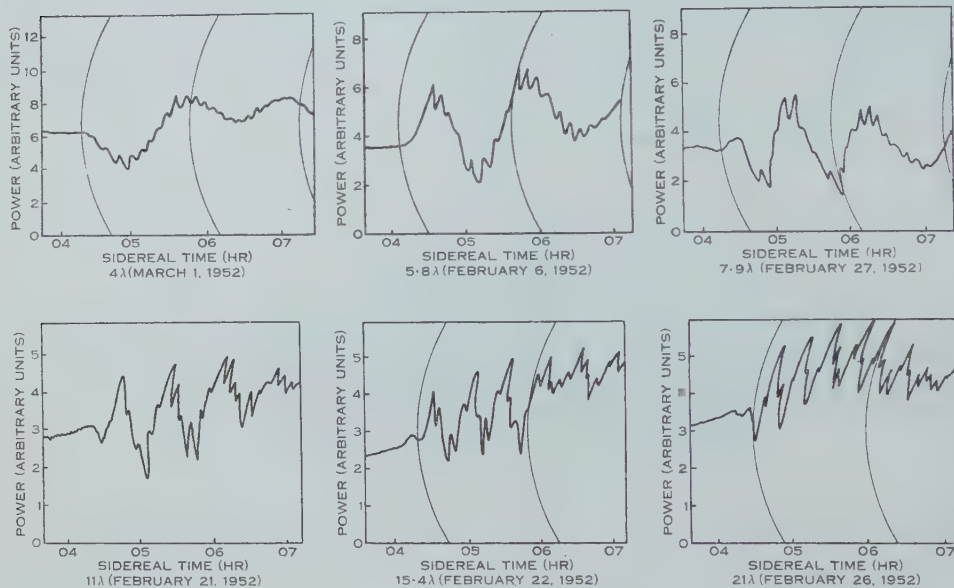


Fig. 3.—Records obtained with the azimuth interferometer (fixed fringes), at six different aerial spacings, of the point and extended sources in Centaurus. The rapid sinusoidal variation is due to the point source passing through the sea-interference fringes, the slower one to the azimuth fringe system. The amplitude of the sea-interference fringes due to the point source provides a calibration. Note that the amplitude of the azimuth pattern due to the extended source increases as the aerial spacing decreases (it is zero for a spacing of 21 wavelengths). The aerials were directed towards an azimuth of 40° E. of S.: the hour angle for a source on the horizon at this azimuth is approximately 08 hr 20 min.

with its central line vertical, the period of the locus does not change. In other cases the period of the locus during rising is shorter or longer than in the remainder of the record, according as the instantaneous centre moves north or south in azimuth relative to the azimuth of a fixed point on the celestial sphere. The change in period of the locus as the source rises would, in principle, enable the inclination of its central line to be determined.

III. OBSERVATIONS OF THE SOURCES

The existence of more than 20 sources has been inferred from the observations with the three aerial systems. Of these, one has been observed with all three equipments, three with two, but the majority with the azimuth inter-

ferometer only. The survey with the 72-ft reflector revealed only those with high flux density within the small region covered. With the sea interferometer and receiver with automatic gain control, sources whose dimensions in the vertical plane at rising are small compared with the aerial beam were discerned. Objects of angular width smaller than 2° could be studied satisfactorily using the azimuth interferometer at aerial spacings of more than 10 wavelengths. With this instrument, as the aerial spacing was reduced, the number of sources observed increased, but so also did the confusion due to the effects of more than

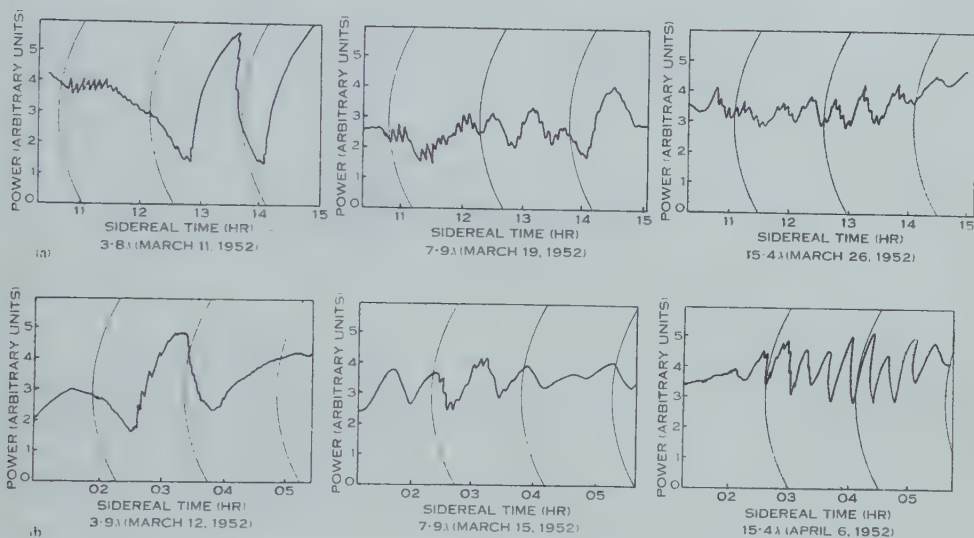


Fig. 4.—Records, illustrating the high degree of confusion of the effects due to a number of sources in the fringe system at the one time, obtained with the azimuth interferometer at three different aerial spacings. (a) A region close to the galactic plane. All records have the same scale of sensitivity, as can be seen from the amplitudes of the sea-interference patterns due to two point sources. The aerals were directed towards an azimuth of 110° ; the hour angle for a source on the horizon at this azimuth is approximately 05 hr 20 min. (b) A region well away from the galactic plane. All the records have the same scale of sensitivity, as can be seen from the amplitudes of the sea-interference patterns due to a point source. The aerals were directed towards an azimuth of 70° ; the hour angle for a source on the horizon at this azimuth is approximately 06 hr 40 min.

one source in the fringe system at a time. For aerial spacings of four wavelengths (fringe separation 15°) interference patterns were observed whenever the plane of the Galaxy crossed the fringe system.

A case where there is little confusion is illustrated in Figure 3 which shows records, obtained with the azimuth interferometer with a stationary fringe system, of an extended source surrounding the "point" source Centaurus-A. At the 21-wavelength spacing only the pattern due to the point source passing through the sea and azimuth fringe systems can be seen. At smaller spacings this pattern is superimposed on the azimuth pattern due to an extended source. The scales of these records are not all the same, but the pattern due to the point source provides a calibration. It can be seen that the amplitude of the

pattern due to the extended source, relative to that of the point source, increases as the aerial spacing is decreased. The azimuth patterns due to the point and extended sources are in phase, suggesting that the two sources are concentric.

Figures 4 (a) and 4 (b) are records, obtained with the same equipment, showing cases with a high degree of confusion, in regions respectively close to and away from the galactic plane. It is not possible to say with any certainty where a source rises on these records. Both the amplitude and period of the pattern at any aerial spacing change in an irregular manner and there is no systematic change in the amplitudes of the patterns for different aerial spacings. Figure 5 is a 24-hr record obtained with the fringe-swinging system. It shows an almost continuous complex pattern due to many extended sources, and sea-interference patterns due to the point sources Taurus-A and Virgo-A.

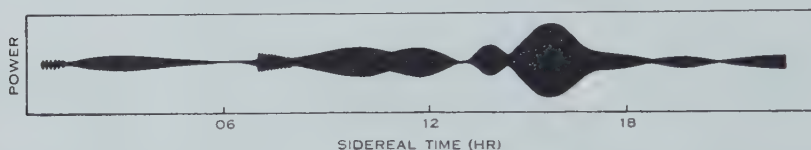


Fig. 5.—A 24-hr record obtained with the azimuth interferometer with the fringe-swinging system. The envelope is filled in by the rapid movement of the recorder pen as the fringe system is swung backwards and forwards. The complex effects due to a number of extended sources passing through the fringe system and the sea interference patterns of two point sources, Taurus-A and Virgo-A, can be seen. The aerials were directed towards an azimuth of 116° ; the hour angle for a source on the horizon at this azimuth is 05 hr.

No attempt has been made to sort out many of the sources from these complex patterns. The accuracy of the results would not warrant the labour involved and the observations must ultimately be superseded by surveys made with pencil-beam aerials.

In Table 1 are listed a number of sources which could be delineated without too much difficulty. The observations on these sources were such that the times of rising above the horizon could be confidently estimated and, with the azimuth interferometer, the changes in amplitude of the patterns with aerial spacing behaved in a regular manner. Table 1 includes the estimated positions, flux densities, and angular extents of the sources, and the equipments with which they were observed. Further details are given in the following notes.

Source A.—This source is clearly seen on the azimuth-interferometer records with aerial spacings less than 15 wavelengths, but its effects are obscured by confusion at smaller spacings. It appears to be the source Fornax-A (Stanley and Slee 1950) and the position given is a new determination from sea-interference measurements.* The declination agrees with that given by Mills (1952*b*) for this source, but the Right Ascension differs by 2 min from Mills's value, a discrepancy which cannot be resolved. Mills also considers that this source has

* These measurements will be reported in a later paper.

TABLE 1
LIST OF SOURCES

LIST OF SOURCES

Source	Approximate Position of Apparent Centroid				Estimate of Flux Density at 100 Mc/s (Both Polarizations) (10^{-24} W m $^{-2}$ (c/s) $^{-1}$)	Estimate of Angular Size (To Approx. 20% of Central Brightness)	Equipment Used in Observations*	Remarks
	Celestial Coordinates (Epoch 1950)		Galactic Coordinates					
	R.A. (hr min)	Dec.	Longitude	Latitude				
A	03 17	-37 $\frac{1}{4}$ °	206°	-56°	6	$\frac{1}{2}$ -1°	Az.I	Probably source Fornax-A
B	04 36	20°	145°	-16°	20	10×5°	A.G.C.	Lies along a parallel of galactic latitude
C	05 10	-43 $\frac{1}{2}$ °	215°	-35°	15	1-2°	Az.I	Possibly source Pictor-A
D	05 42	0°	173°	-14°	15	10×5°	A.G.C.	Lies along a parallel of galactic latitude
E	08 20	-42 $\frac{1}{4}$ °	227°	- 2°	15	1-2°	Az.I	Lies along a parallel of galactic latitude
F	08 24	-44°	230°	- 2°	70	5°	72-ft reflector Az.I, A.G.C.	Source Puppis-A
G	09 15	-10°	210°	+27°	>10	—	Az.I	Position may be in error by several degrees
H	12 40	7°	270°	+69°	25	4°	Az.I	Position may be in error by several degrees
J	13 22	-43°	274°	+20°	50	2°	A.G.C.	Concentric with Centaurus-A
K	16 51	-45°	309°	- 2°	350	10×6°	Az.I	Lies along a parallel of galactic latitude
L	17 41	-27 $\frac{1}{2}$ °	329°	0°	≥300	12×2°	72-ft reflector Az.I	Lies along the galactic equator

* The azimuth interferometer is denoted by Az.I and the sea interferometer with automatic gain control by A.G.C.

an angular diameter of about $\frac{1}{2}^\circ$. The difference in the two values of the Right Ascension may be due to its angular extent, but no definite suggestion can be made as to a brightness distribution which would reconcile the two results.

Sources B and D.—These sources were seen only with the sea interferometer with automatic control of the receiver gain. They are objects of low surface brightness and rise with their central lines nearly parallel to the horizon and the galactic equator. The positions given may be in error by 1 or 2° . The sources were not observed with the azimuth interferometer, presumably because of their extent in azimuth.

Source C.—This source is clearly seen with the azimuth interferometer with aerial spacings greater than 15 wavelengths, but its effects are obscured by confusion at smaller spacings. It may be associated with the source Pictor-A (Stanley and Slee 1950).

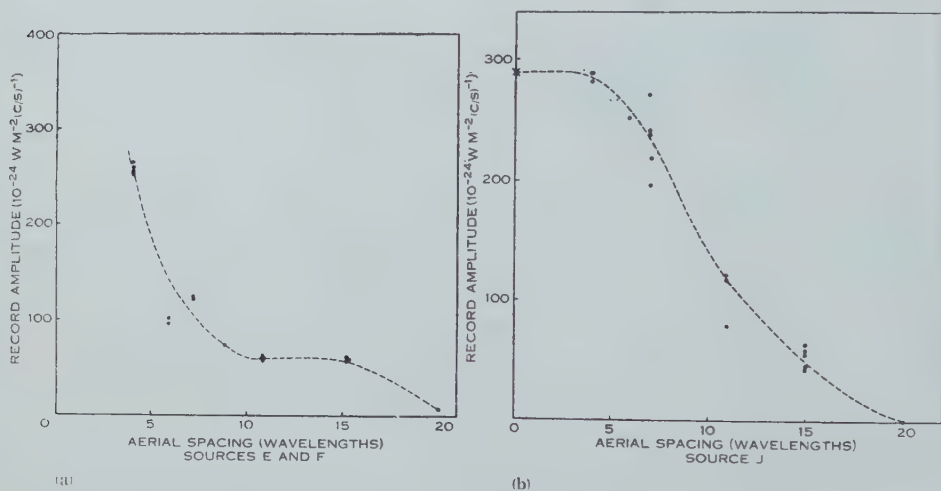


Fig. 6.—Curves showing the variation of the amplitudes of the azimuth-interferometer patterns with aerial spacing. (a) Sources E and F. The effects of source F begin to show for aerial spacings of less than 10 wavelengths. (b) Source J.

Sources E and F.—The observations with the azimuth interferometer in this region were very good. The variations of the amplitude of the interference patterns with aerial spacing are shown in Figure 6 (a). It is believed that for spacings greater than 10 wavelengths the curve is due to source E alone and, at smaller spacings, to the two sources. The position of the centre of source F is difficult to estimate, but from observations with the sea interferometer it appears to be slightly south of E, which is the source Puppis-A (Stanley and Slee 1950). The position of the latter is a new determination.* It disagrees with the position published by Mills (1952*a*, source 08—4) but Mills has informed the authors (personal communication) that his results on this source are ambiguous and that an alternative interpretation would agree with the present results.

* These measurements will be reported in a later paper.

Sources G and H.—These sources are clearly seen with the azimuth interferometer with aerial spacings between 8 and 15 wavelengths but their effects are lost in confusion at smaller spacings. The positions may be in error by several degrees.

Source J.—The azimuth interferometer results in the region of this source are very good and clear of confusion for aerial spacings of more than five wavelengths. The amplitudes of the patterns for various aerial spacings are shown in Figure 6 (b), the value at zero spacing being a direct observation with the sea interferometer. This source overlaps the source Centaurus-A, which has been identified with the galaxy NGC 5128 (Bolton, Stanley, and Slee 1949).

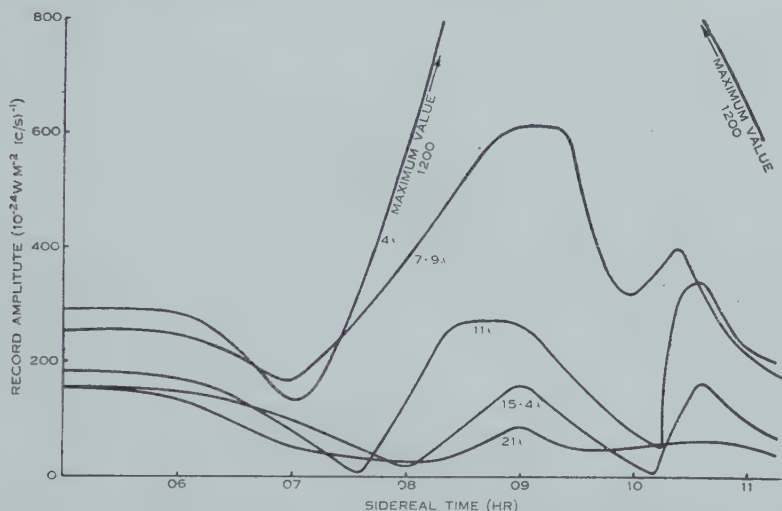


Fig. 7.—Variation in amplitude of the azimuth-interferometer patterns for five aerial spacings. The aerals were directed towards an azimuth of 40° ; the hour angle for a source on the horizon at this azimuth is approximately 08 hr 20 min.

Source K.—This is the second brightest of the extended sources observed. It appears in the contours of the 150 Mc/s survey and is clearly observed at all aerial spacings with the azimuth interferometer. The observations indicate that it rises with its central line nearly vertical, along the parallel of galactic latitude $b = -2^\circ$. It is therefore particularly suitable for analysis of its brightness distribution.

As the source rises, the variation in amplitude of the azimuth-interferometer pattern, for any spacing, determines the distribution along its length of flux density per unit galactic longitude. In the present circumstances, it is approximately given by the rate of change of amplitude with altitude of the leading edge of the source. In Figure 7, the amplitude due to the sources J, K, and L is plotted against sidereal time. At 07 hr, J is passing out of the aerial beam and K is rising into it; at 10 hr, K is passing out of the beam and L is rising into it; at 09 hr, K is fully risen and is making practically the full contribution to the power received. In the interval between about 07 hr 30 min and

08 hr 30 min the curves are almost rectilinear. Thus, for a rough analysis, we may consider the flux density per unit galactic longitude of the source K to be uniform over its extent of approximately 10° .

When fully risen, the variation in amplitude with aerial spacing enables the flux density per unit galactic latitude, in the transverse direction, to be determined by Fourier analysis similar to that used by Stanier for the Sun (see Ryle 1950). Figure 8 (a) is a plot of amplitude against the parameter $2\pi s/\lambda$, where s is the spacing and λ the wavelength; the curve is extrapolated to meet the amplitude axis at right angles. The derived distribution, assumed symmetrical about the central line, is given in Figure 8 (b). The effect of adopting a higher value for the amplitude at $s=0$ would be to raise all the values of flux density per unit latitude, and thus enhance the "skirt" of the distribution.

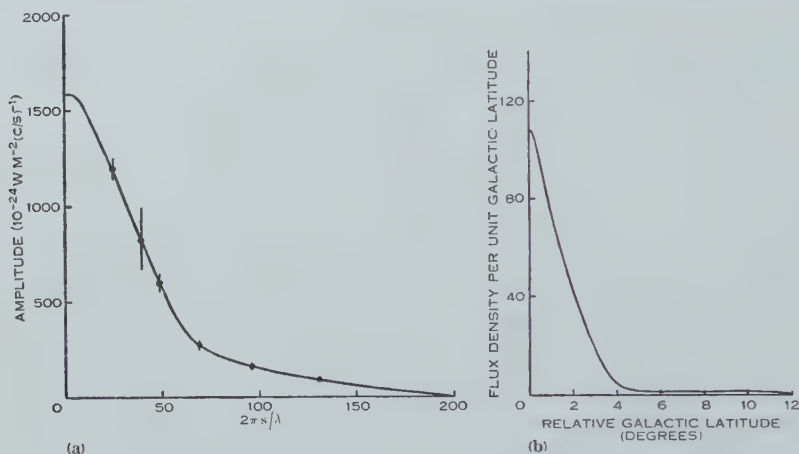


Fig. 8 (a).—Curve representing amplitude as a function of $2\pi s/\lambda$ for the source K. Experimental points are shown as dots, with vertical lines denoting the estimated probable errors. (These decrease with increasing s/λ .) The value at $s=0$ is obtained by extrapolation.

Fig. 8 (b).—The flux density per unit galactic latitude across the source K. Latitude is measured from the central line $b=-2^\circ$. The estimated extent of the source in galactic longitude is 10° , giving a maximum brightness temperature of $12,000^\circ\text{K}$.

However, it is possible that the values of amplitude for the smaller spacings are too high because of contributions from the background. If the background distribution is not symmetrical about $b=-2^\circ$, it will have the effect of broadening the apparent distribution in latitude of the source. Taking the extent in longitude as 10° , we obtain $12,000^\circ\text{K}$ for the mean brightness temperature along the central line of the source.

Source L.—This source, in the region of the galactic centre, is the brightest of the extended sources. Observations with the azimuth interferometer show that it has a strong central concentration, as interference patterns are obtained with the largest aerial spacings (fringe separation 3°). It is not easy to make an analysis of the azimuth-interferometer results as the patterns of this source overlap with those of source K and also the inclination of its axis is unfavourable.

Some idea of its shape may be gained from the results of the 150 Mc/s survey in this region, although the 6° beam of the aerial cannot resolve all the detail.

Observations of the brightest sources K and L were also made with simple azimuth interferometers consisting of two single-Yagi aerials at 100 and 160 Mc/s. It was observed that the ratio of the amplitudes of the interference patterns to the total background noise received by the aerials was approximately the same on the two frequencies, at each of three fringe separations (3, 7, and 14°). These observations suggest that in this frequency range the spectra of the sources are the same as that of the background radiation in their neighbourhoods.

IV. DISCUSSION

The observations described in this paper have revealed the existence of a large number of objects having angular widths of 1° or more. Some of these, for example, sources B, D, K, and L, which are elongated along parallels of galactic latitude, may represent fine structure in the distribution of the background radiation; the remainder appear to form no one distinct physical class. However, sources E and J are of particular interest.

Source E, Puppis-A (08 hr 20 min, $-42^\circ 15'$).—The new determination of position and observation of angular width have led to the identification of this source by Baade and Minkowski (personal communication) with a network of gaseous filaments similar to that which coincides with the Cassiopeia source. The filaments are in an area 1.25 by 0.75° , position angle 135° , centred at 08 hr 20 min, $-42^\circ 48'$. The radio position is slightly different, but this was determined from sea-interference measurements, which refer to a bright concentration rather than the optical centre. Moreover, this position coincides with the most outstanding filament. The filaments show no sign of organized motion, but the velocity dispersion within individual filaments ranges from $+120$ to -30 km/sec. This is the second of a new type of galactic nebulosity discovered through radio observations. The velocity dispersion in the filaments of this source is only about one-tenth, its brightness about one-fifth, but its angular diameter is about 10 times that of the Cassiopeia source. The identification provides another example of a high ratio of radio to optical emission, associated with a violent gas velocity dispersion.

Source J (13 hr 22 min, -43°).—The angular diameter of this source is about $2\frac{1}{2}^\circ$. It overlaps the source Centaurus-A, which has been identified with the galaxy NGC 5128. Centaurus-A is known to have an angular diameter of less than $7'$ (Stanley and Slee 1950; Mills 1952*e*), which is less than the visible extent of the galaxy. It is possible that the small source is associated with the nucleus of the galaxy and the extended object with its outer regions. It would be a remarkable coincidence if these two bright sources had no physical connection.

Of the sources that may be associated with galactic structure, K and L are deserving of further comment.

Source K (16 hr 51 min, -45°).—Observations of this source provide the best example of the type of pattern observed whenever the plane of the Galaxy crosses the fringe system of the azimuth interferometer. It is inferred

that there is a general distribution of similar objects around the galactic plane, which may prove to be fine structure of the background radiation. The analysis of the flux density per unit galactic latitude across source K then indicates a far greater concentration about the plane of the Galaxy than has hitherto been inferred. The half-width deduced from Bolton and Westfold's (1950) 100 Mc/s survey with a low-resolution aerial was 15° , which gave rise to the supposition that the sources of the background radiation were distributed in a manner similar to the stars of Population II. The present value of 3° (see Fig. 8 (b)) is rather to be associated with the objects of Population I, such as early-type stars and interstellar gas and dust. It is possible that the background radiation originates in the interstellar gas, although it is generally agreed that some non-thermal process must be responsible.

Source L (17 hr 41 min, $-27\frac{1}{2}^\circ$).—The observations indicate that this source provides the greatest flux density and has the most peaked brightness distribution in both longitude and latitude of all the extended sources. The position of its centre is close to the accepted position of the galactic centre.

It is difficult to believe that its high flux density is due to the fortuitous superposition of radiation from a number of objects in the line of sight. We are left with the inference that there is an extended physical object at the centre of the Galaxy, which is an unusually intense source of radio noise.

V. LIMITATIONS OF INTERFERENCE TECHNIQUES

The extended sources have been found by the use of interferometers of much smaller aerial spacing, that is, much greater fringe separation, than normally employed. They were not previously discovered in surveys of the general background because of the low resolving powers of the aerial systems used, and, in spite of their high flux densities, were not observed with other interferometers because the fringe separations of those instruments were less than the angular dimensions of the sources. Although the present observations have been of some value, they have also served to emphasize some of the fundamental limitations of interference techniques.

It has been claimed that an interferometer has a resolving power equal to that of a single aerial whose physical dimensions in one direction are equal to the spacing between the individual aerals forming the interferometer. This resolving power can, however, only be realized in the study of a single isolated object. An interferometer consisting of two aerals has a theoretical resolving power determined only by the ratio of the wavelength and the aerial spacing, but in practice it is more often limited by the width of the primary beam of the individual aerial and the distribution of the sources whose angular dimensions are less than the fringe separation within that primary beam.

A source may be regarded as *effectively isolated* under certain circumstances, e.g. when its flux density far exceeds that of any other source within the primary aerial beam; examples are the Sun at centimetre wavelengths and the bright sources in Cygnus and Cassiopeia at short metre wavelengths when the primary aerial beam is fairly small. Two other factors tend to isolate sources in the

case of interferometers with very narrow fringe separations ; firstly, the instrumental effect of a finite receiver bandwidth reduces the visibility of high-order (off-axis) fringes and so effectively decreases the primary beam of the aerial, helping to isolate a source ; secondly, as the fringe separation is reduced below the angular dimensions of most of the sources within the beam, these no longer contribute to the output and leave one source effectively isolated.

Under normal conditions, where there is no isolation, the output of the interferometer represents the sum of the effects of a distribution of sources within the primary beam. If the sources can be considered as widely separated points, the actual distribution can be reconstituted from the observed patterns, provided observations are taken with a number of fringe separations and a number of different interferometer axes. Such reconstitution presents far greater difficulty where the individual sources cannot be considered as points, that is, when their angular dimensions are of the order of the fringe separation and their isophotes are of irregular shape.

The observations described in this paper have shown that the actual distribution of the extended sources is too complex to be delineated by interferometers of moderate primary beam width, since in general there are several of them in the primary aerial beam at a time. It might be argued that some improvement could be gained by reducing the beam width of the individual aerals. This, however, has the effect of reducing the already small number of fringes within the primary beam and, although it might reduce the confusion between a number of sources in the beam at one time, it also introduces the complication of a change in shape from one fringe to the next within the primary beam. It is clear that a detailed study of the extended sources can be made satisfactorily only with pencil-beam aerals whose beam widths are less than the angular dimensions of the sources.

VI. REFERENCES

- BOLTON, J. G., and SLEE, O. B. (1953).—*Aust. J. Phys.* **6** : 420-33.
BOLTON, J. G., STANLEY, G. J., and SLEE, O. B. (1949).—*Nature* **164** : 101.
BOLTON, J. G., and WESTFOLD, K. C. (1950).—*Aust. J. Sci. Res.* **A 3** : 19-33.
BROWN, R. H., and HAZARD, C. (1951a).—*Mon. Not. R. Astr. Soc.* **111** : 576-84.
BROWN, R. H., and HAZARD, C. (1951b).—*Mon. Not. R. Astr. Soc.* **111** : 357-67.
MILLS, B. Y. (1952a).—*Aust. J. Sci. Res.* **A 5** : 266-87.
MILLS, B. Y. (1952b).—*Aust. J. Sci. Res.* **A 5** : 456-63.
MILLS, B. Y. (1952c).—*Nature* **170** : 1063.
PIDDINGTON, J. H., and MINNETT, H. C. (1952).—*Aust. J. Sci. Res.* **A 5** : 17-31.
RYLE, M. (1950).—*Rep. Progr. Phys.* **13** : 184-246.
STANLEY, G. J., and SLEE, O. B. (1950).—*Aust. J. Sci. Res.* **A 3** : 234-50.

GALACTIC RADIATION AT RADIO FREQUENCIES

VIII. DISCRETE SOURCES AT 100 MC/S BETWEEN DECLINATIONS $+50^\circ$ AND -50°

By J. G. BOLTON,* G. J. STANLEY,* and O. B. SLEE*

[*Manuscript received October 19, 1953*]

Summary

One hundred and four discrete sources have been found from a survey covering declinations $+50$ to -50° . The individual sources are compared in position and flux density with those of previous surveys. The observed distribution shows the concentration of sources of all brightnesses to the galactic equator found by Brown and Hazard (1953) and the concentration of the bright sources to the equator found by Mills (1952*a*, 1952*b*, 1952*c*). A new concentration of faint sources is evident in the southern galactic hemisphere. There is strong evidence for departures from a homogeneous isotropic distribution for the sources outside a 20° zone about the galactic equator. Ten identifications between sources and visible objects are suggested; seven of these are between faint sources and extragalactic nebulae of photographic magnitude about 12.5, and three with galactic nebulosities. One of the latter is the expanding shell of Nova Aquila 1918.

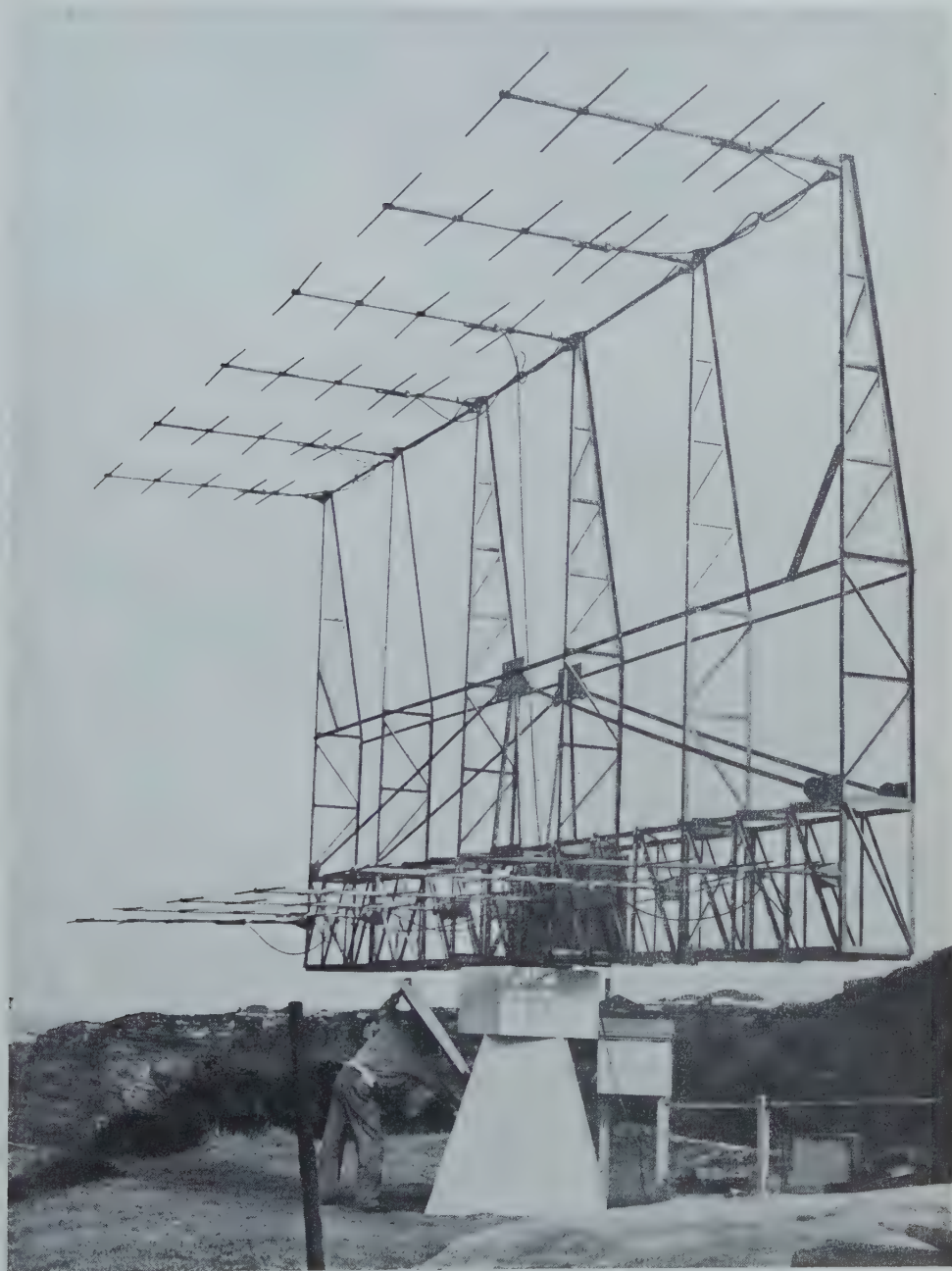
I. INTRODUCTION

This paper describes the results of a new survey at 100 Mc/s of the celestial sphere for discrete sources. It is the sixth major survey; previous surveys have been made by Stanley and Slee (1950), Ryle, Smith, and Elsmore (1950), Mills (1952*a*), Brown and Hazard (1953), and Shain and Higgins (1954). Pertinent details of all the surveys are given in Table 1. The present survey has revealed the existence of the largest number of sources so far, although the number of sources per unit area is somewhat smaller than those of surveys 2 and 4, as is the limit of sensitivity of the survey. The area covered (about 70 per cent. of the celestial sphere) is slightly smaller than that of survey 3 and the limit of sensitivity about the same; the reason for the higher source density is the higher resolving power of the present aerial system.

Sea interference technique was used in the observations, employing the aerial system shown in Plate 1. The aerial diagram in azimuth consists of a main beam approximately 12° between half-power points and a number of very weak side lobes. In the vertical plane the main beam is only about 16° wide due to the large separation between the two banks, but there are major side lobes 20° each side of the main beam. However, interference effects from a source passing through the side lobes do not occur owing to the use of a wide receiver bandwidth and other factors which have been described in a previous paper (Bolton and Slee 1953). Sea reflection reduces the width of the main beam in the vertical plane to about 8° .

* Division of Radiophysics, C.S.I.R.O., University Grounds, Sydney.

GALACTIC RADIATION AT RADIO FREQUENCIES. VIII



12-Yagi array used in the observations. The two vertical banks are about 1.6 wavelengths apart.

The aerial resolving power and the receiver sensitivity are not much greater than those used in the previous survey by two of the authors. The principal reason for the much larger number of sources found in the present survey is the improved receiving technique which permits entirely automatic operation of the equipment and better "seeing" of the interference patterns due to the sources. These improvements have been described by Bolton and Slee (1953).

A comparison has been made between the results of this survey and the others. As has been found by Mills (1952*a*) and Brown and Hazard (1953) in similar comparisons of previous surveys, there is good agreement between the positions and flux densities of the brighter sources. There is also good agreement for some of the fainter sources which are relatively isolated, but the agreement deteriorates for sources in areas of high source density. The reasons for this

TABLE 1
SURVEYS OF DISCRETE SOURCES

No.	Observers	Frequency (Mc/s)	Limit of Sensitivity ($\text{W m}^{-2} (\text{c/s})^{-1}$)	Approximate Aerial Beam (to Half- power Points)	Region Covered (Dec.)	Number of Sources
1	Stanley and Slee (1950)	100	10^{-24}	9 by 17°	50 to -50°	22
2	Ryle, Smith, and Elsmore (1950)	81	3×10^{-25}	$1\frac{1}{2}$ by 90°	90 to 10°	50
3	Mills (1952 <i>a</i> , 1952 <i>b</i> , 1952 <i>c</i>)	100	5×10^{-25}	14 by 24°	50 to -90°	77
4	Brown and Hazard (1953)	158	5×10^{-26}	2 by 2°	70 to 40°	23
5	Shain and Higgins (1954)	18	3×10^{-23}	17 by 17°	10 to -90°	37
6	Present authors	100	5×10^{-25}	8 by 12°	50 to -50°	104

lack of agreement are not difficult to understand. With the interference method of observation, the recorded output of the receiver represents the sum of the effects of all the sources within the primary beam of the aerial. The effects of one outstanding source can be interpreted with a considerable degree of certainty. Where confusion exists on the records, the observer naturally adopts the simplest possible explanation of the complex patterns and trusts that his interpretation is correct. If incorrect, it can result in extreme cases in the assignments of positions and flux densities to sources that do not in fact exist at all. For such sources Minkowski has coined the word "blends". In a particular region of the sky, observations with aeriels of different beams will show confusion in a different manner and degree and the interpretation of the confused patterns may often be different. This situation emphasizes the value of having several surveys with different types of aeriels, for the sources common to the surveys may then be accepted as genuine and not products of individual equipments.

In survey work, the sea interferometer has one advantage over other types of interferometer in that the source suddenly appears over the sharp edge formed

by the Earth's shadow. This feature is of great value in studying regions of high source density. The records obtained in the present survey have been carefully examined for patterns due to sources reported by other observers. The negative results obtained in some cases must, however, be regarded with some caution as there are several possible explanations. The simplest is that the source is a blend and does not exist near the reported position. Another is that the source has a complex brightness distribution and perhaps extends over an angle greater than the fringe separation of the sea interferometer in the direction perpendicular to the axis of the interferometer. Further, in comparing the results of surveys at widely spaced frequencies, it is always possible that, due to an abnormal variation of flux density with frequency, a source which appears relatively bright at one frequency may be too faint to be detected at another.

The 104 sources of this survey have been examined to determine possible trends in the spatial distribution, and 10 identifications between sources and visible objects are suggested.

II. THE OBSERVATIONS

The results of the survey are summarized in Table 2 which consists of five lists of sources. The list to which each source has been assigned depends on the circumstances of its observation, particularly the degree of confusion between the interference pattern of the source and others in its neighbourhood.

List 1 includes a few well-known bright sources, for which accurate positions have been determined and in some cases identifications established with visible objects.

List 2 contains sources for which no confusion exists in our observations and which agree in position with sources reported by other observers.

List 3 contains sources for which no confusion exists in our observations but which have not been listed by other observers. Several of the sources in this list agree approximately in position with those in catalogues by other observers but the agreement is not sufficiently good to warrant their inclusion in list 2.

List 4 contains sources for which some degree of confusion exists in our observations; however, the results given are believed to be fairly accurate.

List 5 contains sources for which a considerable degree of confusion exists in our observations and the particulars given are only reliable provided the observations have been correctly interpreted.

In quite a number of cases the positions of sources in lists 4 and 5 agree well with those of sources reported by other observers.

The first column in these lists contains the authors' personal catalogue number, which places the sources in order of discovery (this system of numbering is continued from the survey by Stanley and Slee (1950)). The second column contains the standard three-letter abbreviation for the constellation in which the

TABLE 2
LIST 1

Authors' Catalogue Number	Constellation	Position		l (deg)	b (deg)	Error in Time of Rising $\Delta(RT)$ (min)	Error in Azimuth at Rising ΔA (deg)	Flux Density S at 100 Mc/s ($10^{-25} \text{ W m}^{-2} (\text{c/s})^{-1}$)	Source Level $\log_{10} S$	Other Catalogue Numbers
		R.A. (hr min)	Dec. (deg)							
10*†	For	03 17	-37½	207	-57	1	2	24	1.4	M03-3
2*††	Tau	05 31	22	152	-4	—	—	185	2.3	M05+2, R05.01
21†§	Pup	08 20	-42½	228	-3	1	2	18 (350)	1.3 (2.5)	M08-4, S08-4
26*	Hya	09 16	-12	212	+26	1	2	28	1.4	M09-1A, S09-1
4*††	Vir	12 28	12½	259	+74	—	—	125	2.1	M12+1, R12.01, S12+1
6*††	Cen	13 22	-42½	277	+19	—	—	180	2.3	M13-4, S13-4
7	Her	16 49	6	352	+27	1	2	40	1.6	M16+0, R16.01
1*††	Cyg	19 58	40½	44	+5	—	—	1200	3.1	M19+4, R19.01, H.B.19, S20+x

* Accurate positions for these sources have been determined by Mills (1952b).

† Mills has shown that these sources have angular diameters of a few minutes of arc (1952c).

‡ Identified with visible objects (Baade and Minkowski 1954).

§ This source has an angular width of the order of 1° . The figures in brackets refer to the total flux density and level, the others to the apparent quantities when the source is observed with the sea interferometer (fringe spacing 1°).|| This source is possibly associated with an object of about 2° angular width close to it (Bolton *et al.* 1954).• The source is possibly associated with objects of larger angular size (Bolton *et al.* 1954).

TABLE 2 (*Continued*)

LIST 2

Authors' Catalogue Number	Position		l (deg)	b (deg)	Error in Time of Rising $\Delta(ET)$ (min)	Error in Azimuth at Rising ΔA (deg)	Flux Density S at 100 Mc/s $(10^{-25} \text{ W m}^{-2} (\text{c/s})^{-1})$	Source Level $\log_{10} S$	Other Catalogue Numbers
	R.A. (hr min)	Dec. (deg)							
71	01 56	-40	222	-70	2	3	5	0.7	M02-4
97	02 36	-3	142	-53	3	5	5	0.7	M02-0
62	02 04	-10	141	-63	2	2	7	0.85	M02-1
40	03 09	41	118	-13	3	4	8	0.9	M03+4, H.B.6 (R03.02)
8	04 30	31	137	-10	2	2	30	1.5	M04+3 (R04.01)
76	05 08	46	130	+5	3	2	7	0.85	H.B.9 (M05+4)
22*	05 09	-43½	215	-35	1	2	25	1.4	M05-4, S05-4
99	08 08	-6	197	+16	2	4	6	0.8	M08-0
86	09 16	46	141	+46	2	4	7	0.85	H.B.12, R09-01 (M09+4)
50	10 20	-43½	245	+12	2	3	8	0.9	M10-4
56	11 38	-15	249	+44	3	3	5	0.7	M11-1
28†	13 35	-60	277	+1	—	—	70	1.8	M13-6, S13-5
80	15 10	11	342	+51	3	2	6	0.8	M15+1
81	16 36	41	31	+41	2	5	8	0.9	M16+4 (R16-03)
27†	16 10	-61	293	-8	—	—	80	1.9	M16-8
120	18 16	-8	350	+2	3	2	15	1.2	M18-0

* The source is possibly associated with objects of larger angular size (Bolton *et al.* 1954).

† The high order fringes of these bright circumpolar sources led to their detection by Stanley and Slee (1950) with the sea interferometer. The positions are those given by Mills (1952a).

Authors' Catalogue Number	Constellation	Position		l (deg)	b (deg)	Error in Time of Rising $\Delta(RT)$ (min)	Error in Azimuth at Rising ΔA (deg)	Flux Density S at 100 Mc/s (10^{-25} W m $^{-2}$ (c/s) $^{-1}$)	Source Level $\log_{10} S$	Other Catalogue Numbers
		R.A. (hr min)	Dec. (deg)							
35	Psc	00 58	15	96	-47	2	3	7	0.85	(M00+1)
38	Cet	00 40	-2	90	-64	2	3	7	0.85	
73	Cet	00 11	-8	70	-70	3	3	6	0.8	
83	Scl	00 23	-29	345	-87	2	4	5	0.7	
92	Tri	01 38	32	104	-29	2	5	6	0.8	(R01.01)
41	Tri	01 30	28	103	-33	2	5	6	0.6	(R01.01)
95	Ari	02 02	15	118	-43	3	3	5	0.7	
53	Cet	03 00	3	143	-44	2	3	9	0.95	
39	Tau	04 06	10	160	-28	2	2	6	0.8	
69	Lep	05 12	-25	194	-30	2	2	7	0.85	
52	Mon	06 26	-6	183	-6	3	2	7	0.85	
70	Pyx	08 57	-25	219	+15	2	3	7	0.85	
96	Leo	09 51	8	199	+45	3	5	7	0.85	(M09+0)
98	Sex	10 42	0	220	+50	3	2	12	1.1	
102	Ant	10 07	-29	234	+22	2	3	7	0.85	
15	Leo	11 42	18	213	+73	2	2	9	0.95	
30	Vir	12 19	5	257	+66	2	2	14	1.1	
65	Com	13 08	30	20	+83	2	3	10	1.0	(M13+2)
111	Hya	13 58	-25	291	+33	2	2	8	0.9	
49	SCt	15 30	20	358	+50	3	5	8	0.9	
79	Boo	14 05	10	322	+63	3	3	7	0.85	
64	SCt	15 58	3	342	+37	2	2	8	0.9	
74	SCt	15 20	-2	329	+41	3	2	5	0.7	
114	Sco	16 44	-43	310	0	2	9	15	1.2	
58	Oph	17 44	-20	336	+2	2	2	20	1.3	(M17-2A)
68	Sco	17 43	-31	327	-3	2	2	12	1.1	(M17-2B)
17	Sco	17 08	-34	320	+2	2	5	15	1.2	
23	Aql	18 45	8	7	+2	4	6	12	1.1	(M19+0)
11	Aql	18 39	2	2	+1	3	2	27	1.4	
48	Aql	19 18	16	18	0	2	2	20	1.3	
75	Sag	19 33	-17	350	-19	3	3	8	0.9	(S19-2)

TABLE 2 (Continued)

LIST 5

Authors' Catalogue Number	Position		l (deg)	b (deg)	Error in Time of Rising $\Delta(RT)$ (min)	Error in Azimuth at Rising ΔA (deg)	Flux Density S at 100 Mc/s (10^{-25} W m ⁻² (c/s) ⁻¹)	Source Level $\log_{10} S$	Other Catalogue Numbers
	R.A. (hr min)	Dec. (deg)							
109	01 35	-12	131	-70	4	10	4	0.6	
47	03 08	-16	169	-54	3	5	6	0.8	
122	04 46	-30	198	-37	5	5	6	0.8	(M04-3)
121	05 48	-17	190	-19	4	2	7	0.85	(M06-1) (S06-1)
32	07 00	15	169	+11	4	10	6	0.8	
93	09 56	28	170	+54	3	3	7	0.85	
19	09 40	0	206	+39	3	5	4	0.6	
119	09 08	-42	234	+5	1	2	10	1.0	
101	10 46	-20	237	+35	4	10	5	0.7	
123	11 00	-32	247	+26	3	7	5	0.7	(M10-3)
78	13 26	10	303	+69	4	5	4	0.6	
110	14 54	-29	303	+25	4	14	10	1.0	
112	15 48	-34	309	+14	3	3	14	1.2	
51	17 12	10	359	+25	5	10	15	1.2	
36	18 47	12	12	+4	3	5	10	1.0	
31	21 13	20	38	-20	5	6	10	1.0	
117	21 30	-42	327	-49	3	20	5	0.7	(M21-4)
45	22 24	4	38	-44	3	4	5	0.7	
115	23 32	-29	352	-75	4	10	5	0.7	(S23-2)

source is situated, the third to sixth the position, and the seventh and eighth the limits of error in the position of the source. The ninth column contains the flux density of the source at 100 Mc/s; the errors in the flux densities relative to that of the source Virgo-A which was taken as the standard are probably not more than ± 20 per cent. Column 10 is the "level" or radio magnitude defined, following Mills, as $L = \log_{10} S$ where S is the flux density in units of $10^{-25} \text{ W m}^{-2} (\text{c/s})^{-1}$. In column 11 are the catalogue numbers assigned to various sources by other observers; where there is some doubt about the identifications the catalogue numbers are in brackets. The following abbreviations are used: M for Mills, R for Ryle, Smith, and Elsmore, H.B. for Brown and Hazard, and S for Shain and Higgins. Additional information on some sources is contained in the footnotes following Table 2.

The limits of error in position are not, as is customary, given in Right Ascension and declination since the factors determining the position are the sidereal time at which the source crosses the optical horizon and the azimuth bearing of the source at that time. For sources of declination greater than about 25° , the rate at which the source passes through the fringe system of the interferometer is used to supplement the latter. The accuracy in the determination of the rising time, $\Delta(RT)$, depends partly on the number of observations of the source, which helps to reduce the uncertainty introduced by the variable atmospheric refraction at low angles and by the degree of confusion. The accuracy in azimuth at rising, ΔA , depends on the number of observations of the source, the degree of confusion, and, in the case of sources near the galactic plane, on the change in the background noise during the observation of the source. (The change in sensitivity of the equipment due to the change in background level effectively alters the polar diagram of the aerial as far as the observation of the source is concerned and thus complicates the direction finding.) The area contained by the limits of error is a parallelepiped which changes shape with the declination (for examples see Stanley and Slee (1950)); its smallest side is generally that determined by the error in time of rising. These errors increase on the average for the sources in the higher list number and in the case of a source in list 4 or list 5 are subject to the interpretation of the observations being correct.

In general nothing certain is known about the angular widths of the sources of this survey. The sea interferometer used in the observations has a fringe separation of 1° and thus it can be assumed that in most cases the angular sizes of the sources are less than 1° . An exception is that of Puppis-A (source 21, list 1) which is known to have an angular width of about 1° overall; however, a sea interference pattern is produced by a bright local patch near one edge of the source (see Bolton *et al.* 1954). It is thought unlikely that any of the sources away from the galactic plane have angular sizes of more than a few minutes of arc, but it is possible that the sea interference patterns of some of the sources near the galactic equator may be due to bright patches or central concentrations of much more extended objects.

III. A COMPARISON OF THE VARIOUS SURVEYS

(a) Overall Comparison

An examination of the sources of the present survey and those given in surveys by Ryle, Smith, and Elsmore (1950), Mills (1952*a*, 1952*b*, 1952*c*), and Brown and Hazard (1953) shows that 3 sources are common to all the surveys, 7 common to three (5 with Mills and Ryle, 1 with Mills and Brown, and 1 with Brown and Ryle), and 35 to two (33 with Mills and 2 with Ryle).

(b) Comparison with Mills's Survey

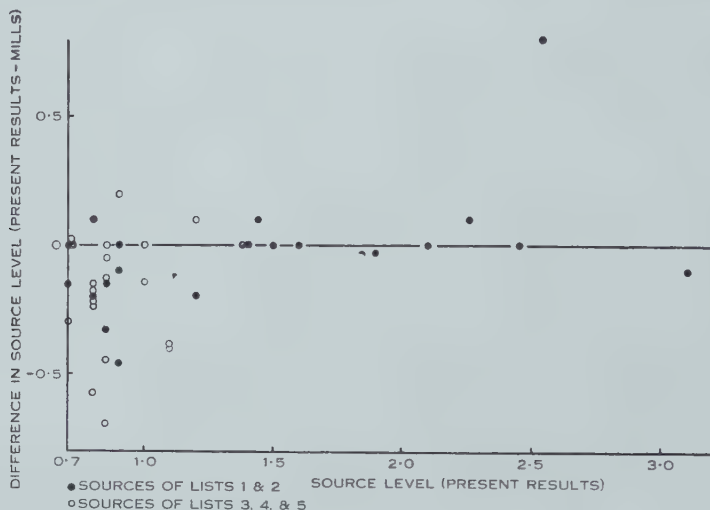
Mills's list contains 69 sources (including the two circumpolar sources 28 and 27) in the area common to his survey and our own. Of these, 42 or 60 per cent. agree in position to within 3° . For these sources the two determinations of the flux densities are compared in Figure 1 (*a*). The abscissa represents the source level of the present survey and the ordinate the difference in levels (i.e. the logarithm of the ratio of flux densities). The dots refer to the sources of our lists 1 and 2 and the open circles to the remaining sources. For the majority of the sources the level differences are less than 0.2 and a scatter of this order may be considered reasonable in view of the experimental accuracy claimed in each case. There is a slight tendency for the level difference to increase for the fainter sources which means that Mills's estimates of the flux densities of the fainter sources are higher than ours.

Nine sources lie well outside the 0.2 scatter in level differences. These sources are: 21, M08-4; 58, M17-2A; 68, M17-2B; 23, M19+0; 123, M10-3; 86, M09+4; 82, M10+4; 76, M05+4; and 40, M03+4. The first, source 21 (Puppis-A), is known to have an angular size of the order of 1° ; our measurements refer to the total flux density and Mills's only to the effective flux density of that part of the central concentration which gives rise to the pattern on his interferometer. The discrepancy is thus resolved in this case. The next three sources are close to the galactic plane; the discrepancies in these cases may be due to the angular sizes being comparable with the fringe separations of the interferometers and the different axes of the interferometers, or may be due to errors in calibration (the sources are in regions of high background radiation). No reason can be offered for the discrepancy in the case of the fifth source except that the error in one or both cases may be due to confusion with neighbouring sources. The four remaining sources have one feature in common—they are all north of Dec. $+40^\circ$. Examination of the level differences for all the sources north of $+40^\circ$ shows that Mills's estimate of flux density is higher than ours in all cases. This suggests a systematic error and Mills (personal communication) has informed the authors that he may have underestimated the effect of ground reinforcement on his results for the sources of high northerly declination.

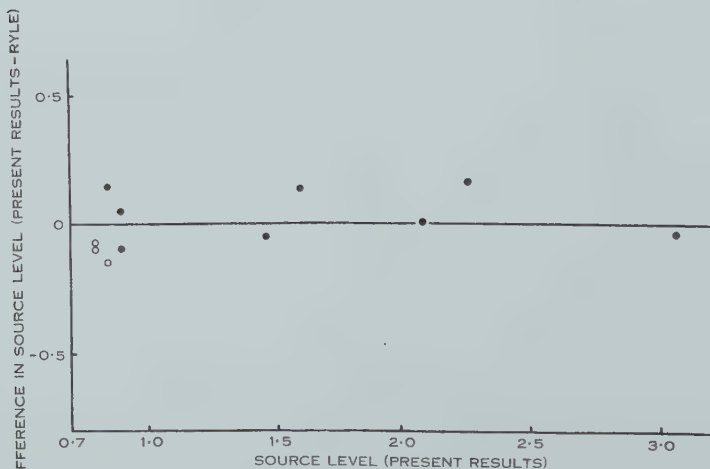
(c) Comparison with Ryle's Survey

In the area common to the two surveys there are 32 sources in our list and 35 in Ryle's list. Only 11 of these appear to be identical, which at first sight seems very poor agreement. However, of the 32 in our list, many are of low

declination where the sensitivity of Ryle's interferometer is fairly low. Further, of the 35 in his list, many are below the limit of detection of our instrument.



(a)



(b)

Fig. 1 (a).—Comparison between the “levels” of sources common to the present survey and that of Mills. The abscissa is the level given by the present survey and the ordinate the difference between the present value and that due to Mills. The dots refer to sources in lists 1 and 2 and the open circles to the remainder.

Fig. 1 (b).—A similar comparison between the sources common to the present survey and that by Ryle, Smith, and Elsmore.

Of the 17 sources within the limit of detection, 11 are identical with sources in our list and thus the agreement is of about the same standard as in the previous comparison with Mills's results.

The level differences are compared, as before, in Figure 1 (*b*). There is less scatter in the level differences than in the previous comparison and no change with absolute levels. This comparison would indicate that there is no change in flux density with frequency. Observations by Stanley and Slee (1950) of the radio-frequency spectra of four bright sources and other comparisons of individual measurements by Shain (1954) and Brown and Hazard (1953) suggest spectra of the form $S \propto f^{-n}$ where n is of the order of unity. The difference in levels should therefore be $n \log(f_1/f_2)$ which is -0.1 for n unity and f_1 and f_2 , 100 and 80 Mc/s. The fact that this difference does not show up in Figure 1 (*b*) suggests a systematic error in the measurement of flux densities in one or both surveys.

(d) Comparison with Brown's Survey

In the area common to the two surveys there are eight sources in our list and nine of flux density greater than $3 \times 10^{-25} \text{ W m}^{-2} (\text{c/s})^{-1}$ at 158 Mc/s in Brown's list. Of these nine, two are extended objects and one is on the northern limit of our survey. Of the remaining six, five agree well in position. There are so few identical sources that the levels are not worth comparing in detail. However, it appears that the flux density at 100 Mc/s is greater in all cases than that at 158 Mc/s but there is a considerable range of actual ratios.

(e) Comparison with Shain and Higgins's Survey

There are 74 sources in our list and 31 in Shain and Higgins's list (including the circumpolar sources 27 and 28). Of these, 13 are common to the two—a somewhat poorer result than the previous comparisons. However, it should be remembered that the observing frequencies are much further apart in this case and that Shain and Higgins's method of observation is very different from ours. They did not employ an interferometer but deduced the existence of the sources from local "peaks" on the contours of the background distribution at 18 Mc/s, and they point out that some of these sources may therefore have large angular sizes, well in excess of the fringe separation of our interferometer. We have not considered the difference in levels for the sources common to the two surveys as Shain (1954) has already made a detailed comparison between his and Mills's results.

(f) Reliability of the Present Survey

These comparisons enable us to make some estimate of the general reliability of the results of the present survey. The agreement between the individual sources of the various surveys has been found to be quite high after due allowance has been made for factors such as the different sensitivities of the instruments and conditions of the surveys. Sources which may be considered as definitely existing are those which can be identified with sources reported by other observers—the 8 sources of list 1, the 16 sources of list 2, 8 of the sources of list 4, 5 of the sources of list 5—a total of 37. To these we may confidently add the 36 sources of list 3 (8 of these are possible identifications with sources of other observers) giving a total of 73 or 70 per cent. of the total. It is also reasonable to assume that quite a number of the remainder in lists 4 and 5 are correctly reported so that the percentage is probably much higher.

IV. SUGGESTED IDENTIFICATIONS

The ultimate purpose of any survey such as this is the "identification" of the sources. This can be done in two ways; the direct method of correlating the positions of radio sources with visible objects and the indirect method of studying the distribution of the sources. The latter may lead to identification through the finding of radio and visible emitters in a common spatial distribution.

As far as the first approach is concerned, we may rely on the already established identifications between sources and visible objects to guide the search. The choice seems to lie between extragalactic nebulae with certain anomalies and galactic nebulae whose common characteristic is a large dispersion of gas velocities. The latter type is mostly objects of extremely low luminosity and present some difficulties to the astronomer with quite large telescopes. The radio astronomer, armed only with a collection of star atlases or catalogues, can hardly expect to suggest identifications of this type. With extragalactic nebulae the situation is rather different; the difficulty of the low luminosities does not always exist but a difficulty of position accuracy is immediately apparent. The positions of the sources are only approximate, the average limit-of-error rectangle for the sources of the present survey covers about 5 deg^2 and there are about 100 sources, or one to every 300 deg^2 of the sky covered. In the same region, reference to the Shapley-Ames catalogue of extragalactic nebulae shows that the average density of galaxies down to the 13th magnitude is about $1/30 \text{ deg}^2$. Thus there appears to be a one-in-six chance of finding a galaxy brighter than 13th magnitude within the limit-of-error rectangle of a source. For 14th magnitude galaxies the chances are nearly even. It is apparent that suggesting identifications on such an arbitrary basis without some supporting evidence would be quite unjustified.

However, a suggested identification, if it can be made, is valuable for two reasons. Firstly, it draws to the attention of the astronomer the possibility of peculiarities in a galaxy about which practically nothing may be known except its position in the sky and a very rough description of its appearance. Secondly, it gives hope to the radio astronomer, intent on making accurate determinations of position, that a certain identification may result from his efforts—for there are cases where an extremely accurate position has not led to the identification of a source with any visible object. These cases represent a certain waste of effort.

A more positive approach to the question can be made if, in addition to position measurements, we take account of the ratio of optical and radio emission of the already established extragalactic radio emitters. At one end of the scale Brown has shown that there is a fairly uniform relation between the optical and radio emission of certain (assumed) normal galaxies including our own Galaxy and several others of the local group of galaxies. Of these only the flux density of Andromeda (and the Magellan Clouds possibly) is sufficient to enable their detection with the present equipment. The angular sizes are, however, too great for detection with the interferometer of fringe separation of 1° . At the other end of the scale there is the second brightest radio source, that in Cygnus, which is identified with a pair of very faint galaxies in collision. With such

a high ratio of radio to optical emission, it would be hopeless to search for objects with a similar ratio whose radio flux density is 100 times smaller. In between, there are two and possibly three galaxies whose ratio of radio to optical emission is about 100 times that for the normal galaxies. These are the galaxies NGC 4486, 5128, and possibly 1316. To find the level difference, which is the logarithm of the ratio of optical to radio intensities in common units, we have to multiply the optical magnitude by 0.4 and add it to the radio level. This gives level differences in the three cases of 6.3, 5.1, and 5.4; these are fairly close and have a mean of 5.7. If there are other galaxies with the same ratio, for sources down to a level of 0.8, they should be brighter than 13th magnitude. On this basis we suggest the seven galaxies in Table 3 as possible abnormal radio

TABLE 3
SUGGESTED IDENTIFICATIONS OF SOURCES WITH EXTRAGALACTIC NEBULAE

Source	List	Galaxy			Remarks on Position Agreement	Ratio of Radio to Optical Level, $L+0.4m$
		NGC Number	Type	Mag.		
38	3	227	—	13.1	Observations and position agreement very good	6.1
102	3	3125	—	13.0	Some scatter in observations. Positions within 1°	6.0
30	3	4303	SBe	10.4	Observations and position agreement very good	5.3
88	4	6482	E	12.2	Some scatter in observations but line of rising passes through nebula	5.6
77	4	7541	Sb	12.8	Within limits of error	5.7
47	5	1209	—	12.5	Observations fair. Positions within 1°	5.8
19	5	2967	S	12.4	A faint source. Agreement good	5.6

emitters. The radio and optical positions are in good agreement and the difference in levels of the order indicated. Moreover, the galaxies are all in regions of fairly low density (i.e. down to the magnitude limit taken) so that there are no alternative identifications possible from nearby galaxies. The density of galaxies in the regions is far less than the average of $1/30 \text{ deg}^2$, considered earlier, as most of the galaxies down to 13th magnitude are concentrated in the Virgo and Fornax clusters. If a fair proportion of the latter were abnormal, we would not expect to distinguish many of them owing to the resultant confusion in the radio observations.

The average difference in levels for the three identifications and the seven suggested is about 5.7. This may be compared with Brown's value of 3.6 (in equivalent units) for normal galaxies at 158 Mc/s. The correction for the

radio spectrum between 100 and 158 Mc/s is probably not more than 0.4 (assuming a variation of flux density inversely proportional to frequency). Thus the radio emission from the abnormal galaxy exceeds that of a normal galaxy of the same optical magnitude by a factor of about 100. It is suggested that the one exception to Brown's almost uniform ratio for normal galaxies, NGC 891, may also be of this abnormal type. Brown has explained this case by assuming that the dark band of the spiral, which is seen almost edge-on, obscures most of the visible light. However, it seems that the discrepancy is much too great to admit this explanation.

In addition to the extragalactic nebulae, the following galactic objects are suggested as possible radio sources on the basis of the position measurements: source 11 with the expanding shell of Nova Aquila 1918 and sources 58 and 119 with the planetary nebulae NGC 6445 and 2792. In the case of source 11, the radio position is 2° away from the nova shell on the line of rising passing through the nova position. Although 2° is fairly close to the limit of error, it is not large in view of the particular difficulties in this case.* The expanding shell does seem to be the type of object that could produce high radio emission, considering the other objects in which high gas velocities are associated with abnormal radio emission.

V. THE DISTRIBUTION OF THE SOURCES

(a) *Instrumental Effects Affecting the Observed Distribution*

Before examining the distribution of the sources in detail, it is necessary to consider the consequences of instrumental selection on the results. Instrumental selection can be of two types, one due to a change of sensitivity in the equipment with position on the celestial sphere and the other due to a combination of the actual source distribution and the resolving power of the aerial. The first type may be again subdivided.

Firstly, some variation in sensitivity might be expected with the declination of the source. A source of declination 0° passes through the fringe system at the rate of one fringe in 5 min, a source of 40° declination at one fringe in 10 min, and the rate decreases rapidly towards the limit of the survey. Variations in receiver gain, the presence of external interference, and other factors are more likely to affect the "seeing" of the slow interference patterns due to sources of high declination than the rapid ones of low declination. A test was therefore made of the observed distribution to see whether this effect was present in the results. No marked effect was found as can be seen from Figure 2 (a) which shows the distribution of sources amongst seven equal area zones bounded by parallels of declination. The result is compared with the data of Mills's survey in Figure 2 (b). It appears that Mills's survey tended to favour sources of high declination as his results show a relative excess in zone 7, but this may be in part

* The line of rising can be determined with relative certainty but in this case it is difficult to secure an intersecting line and so establish the position of the source on the line of rising. As the declination is very low the method of finding it from the fringe period is inaccurate and, as the source is close to the galactic plane, the change in sensitivity during the period of observation renders direction finding on the source unreliable.

due to the higher sensitivity in high northern declinations already mentioned. Both results show a relative deficiency in zone 6 (declinations $33\frac{1}{2}$ – $41\frac{1}{2}^\circ$).

Secondly, the sensitivity of the equipment varies with the background noise from the Galaxy received by the aerial. The effect of the background noise is inherent in any observations of this type but the extent of the effect is determined by the aerial system and the receiver noise factor. The present aerial with its relatively low resolving power does not reflect the true background distribution, the variation in which is of the order of 20 or 30 : 1. It receives power from all directions within the beam and the variation in this over the celestial sphere is only of the order of 4 or 5 : 1. Moreover, the total output of the receiver is due to the sum of the power received by the aerial and the noise

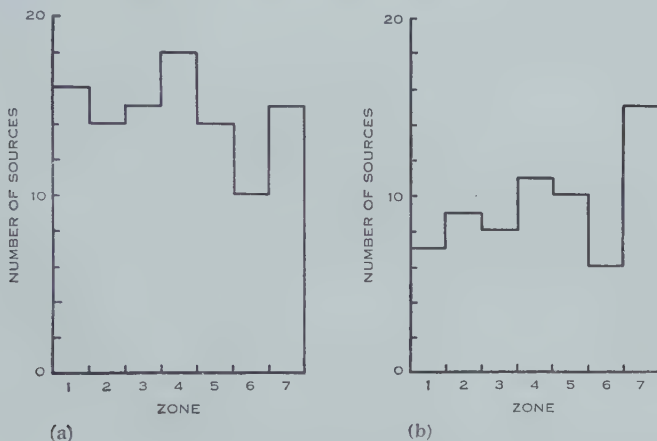


Fig. 2.—Diagrams illustrating the uniformity of the material of the present survey and that by Mills. The number of sources observed in seven equal area zones bounded by parallels of declination north and south of the celestial equator are shown. Zones are ± 0 – $6\frac{1}{2}^\circ$, $6\frac{1}{2}$ – $12\frac{1}{2}^\circ$, $12\frac{1}{2}$ – $19\frac{1}{2}^\circ$, $19\frac{1}{2}$ – $26\frac{1}{2}^\circ$, $26\frac{1}{2}$ – $33\frac{1}{2}^\circ$, $33\frac{1}{2}$ – $41\frac{1}{2}^\circ$, and $41\frac{1}{2}$ – 50° .

generated in the receiver and associated cables. The sensitivity of the equipment is inversely proportional to the total receiver output, and in the present case varies by about 2 : 1 between the aerials directed towards the galactic poles and in the direction of the galactic centre. Undoubtedly the variation in sensitivity* due to this cause is present in the results of our observations but we consider that it is probably overshadowed by another effect in regions near the galactic plane—the shielding due to a concentration of bright sources.

Shielding due to a bright source or a number of bright sources is probably the prominent instrumental effect in the observed distribution of the sources.

* Taking these considerations to their limit of absurdity, the most uniform results from a survey would be obtained for no background variation. This would require either an aerial with no directivity, a receiver with an infinite noise factor, or a frequency of observation where the background noise is zero. All these factors are opposite to those desirable for obtaining large numbers of sources from a survey.

In the neighbourhood of a bright source, fainter sources cannot be detected owing to the overriding effect of the bright one. The shielding by a bright source of the area round it depends on the shape and size of the aerial beam so that different surveys suffer to different extents. In most cases the shielded area extends all round the bright source but with the sea interferometer the pattern due to a faint source can be distinguished until the bright one appears over the horizon. Thus, with a given aerial, observations over the sea contain only half the shielded area of any other type of observation with the same aerial.

(b) *Examination of the Observed Distribution on a Number Basis*

The observed distribution of the sources of this survey is shown in galactic coordinates on an equal-area chart in Figure 3. The brighter sources (with flux densities of greater than $10^{-24} \text{ W m}^{-2} (\text{c/s})^{-1}$) are shown as black dots and the

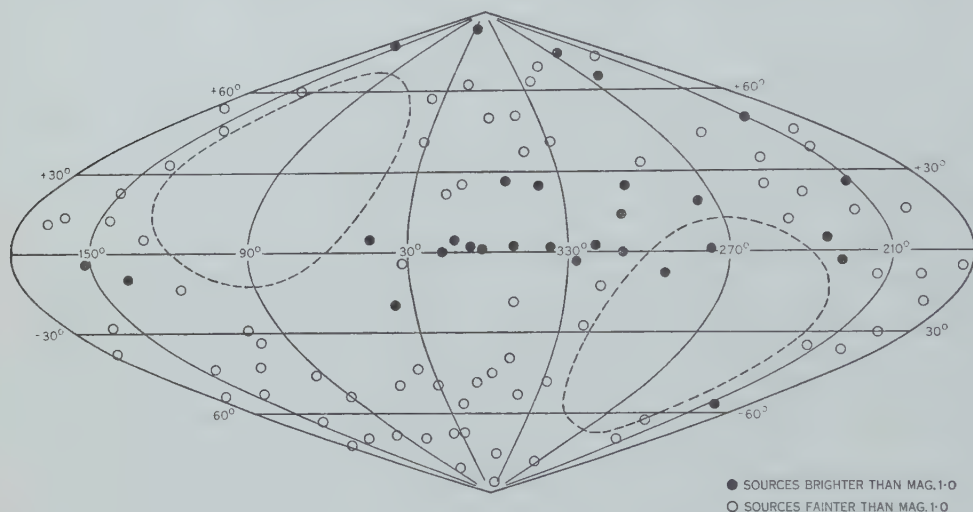


Fig. 3.—The distribution of the sources shown on an equal-area chart in galactic coordinates. The areas enclosed by the dotted lines are north of Dec. 50° and south of Dec. -50° . The black dots are sources whose flux densities are greater than or equal to $10^{-24} \text{ W m}^{-2} (\text{c/s})^{-1}$ at 100 Mc/s (i.e. levels greater than or equal to 1.0).

fainter ones as open circles. The dotted lines enclose the circumpolar regions; the two sources within the south circumpolar area are numbers 27 and 28 which were detected from their high-order fringes (see footnotes to Table 2). Two features stand out in the distribution, the concentration of bright sources to the galactic plane and the concentration of faint sources in the southern polar cap. These concentrations are shown up in Figure 4 (a) which gives the number of sources in each of seven equal-area zones bounded by parallels of galactic latitude on each side of the galactic equator. The black rectangles represent the brighter sources. Some correction is necessary to this figure as some of the zones contain different amounts of the circumpolar areas. Correction on an area basis results in Figure 4 (b) in which the features of Figure 4 (a) are still present in the same

marked manner. (This correction can only be done on a number basis and not magnitude and number.) Figure 4 (c) is the result for the sources of Mills's survey treated in a similar manner. Our results confirm Mills's finding of a concentration of bright sources to the galactic plane but give it as a concentration both in *number* and *magnitude*. The south polar concentration of faint sources is a new result but should be considered in the light of shielding effects. Figure 4 (a) could be considered as showing a relative deficiency of faint sources in low southern latitudes and all northern latitudes. The former might be due to shielding by the effects of the line of bright sources along the galactic equator and the latter by the few bright sources scattered in the northern hemisphere. Shielding in the northern hemisphere could produce the effect observed, but we consider that the relative concentration near the south pole or the relative deficiency of sources in low southern latitudes is too great to be a purely instrumental effect.

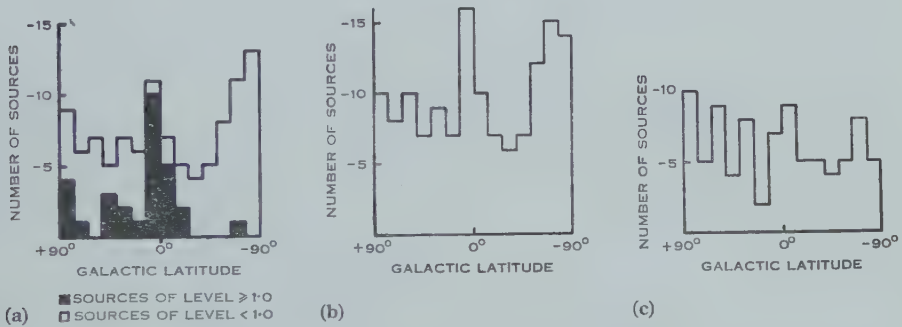


Fig. 4.—The observed distribution of the sources in 14 equal-area zones bounded by parallels of galactic latitude. (a) Showing the distribution of the bright and faint sources; (b) as in (a) but corrected for the differences in those parts of the zones in circumpolar regions; (c) as in (b) for the sources in Mills's list.

(c) Examination of the Observed Distribution in both Number and Level

Both Ryle, Smith, and Elsmore (1950), and Mills (1952a) have investigated the uniformity in the distribution of the sources by considering the relation between the number of sources greater than a certain level and that level. The method is similar to that used in testing the spatial uniformity of extragalactic nebulae. It can easily be shown that for a homogeneous isotropic distribution of sources of the same absolute magnitude (and for a small dispersion in absolute magnitudes) the logarithm of the number of sources greater than a level L is proportional to $-1.5L$ (i.e. $N \propto S^{-3/2}$). Ryle, examining the distribution of some 50 sources, mainly well away from the galactic plane, found just this factor. Mills for a more even sample of the celestial sphere found a factor closer to unity. However, when he divided the sources into those within a zone 24° wide about the galactic equator and those outside, he found a factor of -0.75 for the former and -1.5 for the latter. He suggested that for those near the galactic plane, the result pointed to a distribution in a flattened disk and for those outside an isotropic distribution.

We have followed Mills in analysing our results. Various ogives of the log (number of sources greater than level L) against L are shown in Figures 5 (a)–(e). Figure 5 (a) is for all the sources, (b) for those within 10° of the galactic plane, (c) for the sources outside this zone, and (d) and (e) for the individual hemispheres outside these zones. The last four are in terms of the number of sources per steradian. For all the sources together the slope of the line is somewhat irregular but has a mean of about unity, in agreement with Mills. For the sources within 10° of the galactic plane the slope is about 0.6, again in agreement with Mills. In the other cases the slopes are much steeper than that required by the isotropic distribution. All the curves show some falling off above a level of 0.75 (near the limit of the survey) which is to be expected. The result for the individual hemispheres outside the equatorial zone means that there are more *faint* sources

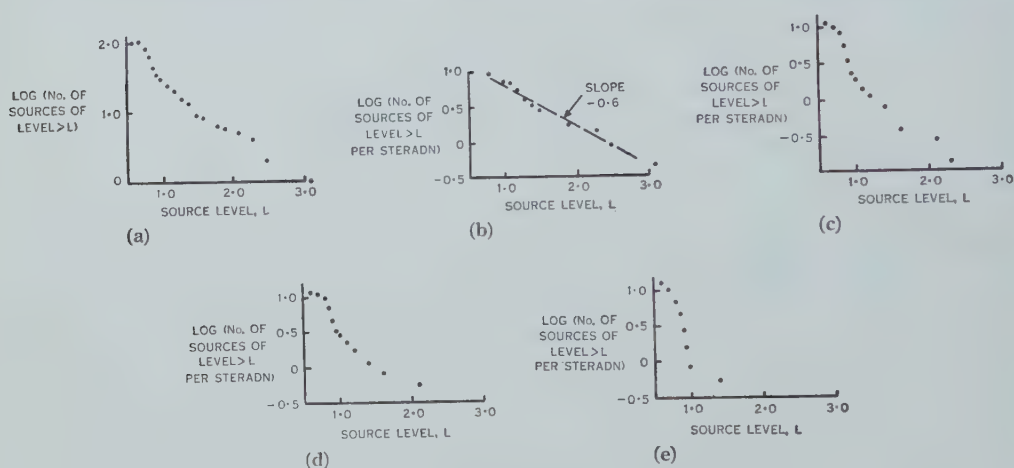


Fig. 5.—Ogives showing the relation between the number of sources greater than a certain level and that level. (a) 104 sources of the survey; (b) 21 sources within a zone 20° wide about the galactic equator; (c) 83 sources outside the zone of (b); (d) 40 sources north of the equatorial zone; (e) 43 sources south of the equatorial zone. In (b)–(e) the ordinates are in terms of the number of sources per steradian.

than would fit in with the idea of the uniform distribution. Correction of the results for the effects of shielding would enhance this trend in the distribution rather than destroy it. Our survey and that of Mills have both approximately the same sensitivity limit; it is clear that with the higher aerial resolving power the additional sources that have been discerned are mainly faint sources and this has made the difference in the results for the area outside the zone about the galactic equator.

VI. DISCUSSION

The results of this survey have (1) offered the possibility of some further identifications of sources with visible objects, principally extragalactic nebulae, and (2) raised interesting points on the distribution of the sources. The latter would appear to warrant some further discussion.

Firstly, the distribution indicates, as Mills has suggested, a concentration of the bright sources to the galactic plane and, as Brown and Hazard have found, an overall concentration of the sources towards low galactic latitudes. The concentration of bright sources shown by either Figure 3 or Figure 4 (*a*) is relatively independent of the flux density limit arbitrarily chosen ($10^{-24} \text{ W m}^{-2} (\text{c/s})^{-1}$); of the 30 brightest sources, 14 are within 5° of the galactic equator. If the limit is made 1.5 units, 11 out of the 18 are within 5° of the galactic equator; if it is 2.0 units 8 out of the 15 sources are within this zone. The overall concentration shows up in Figure 4 (*b*) where the number of sources only is considered and, in addition, it must be remembered that correction for shielding due to the presence of a large proportion of relatively bright sources would have the effect of increasing the total number in the zone about the equator. Of the sources in this zone it is also noticeable that most of them are round the direction of the galactic centre; there are nine sources within $\pm 50^\circ$ of longitude 330° compared with only three within $\pm 50^\circ$ of longitude 150° . The concentration of the sources may thus be twofold, towards the equator and more particularly towards the direction of the galactic centre.

There is also an apparent concentration of sources near the south galactic pole, which does not appear to be due to instrumental effects. For both galactic hemispheres there seem to be too many faint sources compared with an isotropic distribution of objects of all about the same absolute brightness; this again is not believed due to instrumental selection for correction for this would increase the trend of the results. A plausible explanation is that the Sun (if these sources are galactic) or the Galaxy (if the sources are extragalactic) is in a local region of low source density and that somewhere towards the limit of the survey we reach a region of much higher density. Such deviations from uniformity are frequently observed in the statistics of the distribution of extragalactic nebulae—where objects in a large cluster influence the results at a certain stage. However, there is not much point in speculating too far on this result as it could also be produced by a large dispersion in absolute magnitudes amongst the sources of the survey.

VII. REFERENCES

- BAADE, W., and MINKOWSKI, R. (1954).—(in press.)
 BOLTON, J. G., and SLEE, O. B. (1953).—*Aust. J. Phys.* **6**: 420.
 BOLTON, J. G., WESTFOLD, K. C., STANLEY, G. J., and SLEE, O. B. (1954).—*Aust. J. Phys.* **7**: 96.
 BROWN, R. H., and HAZARD, C. (1953).—*Mon. Not. R. Astr. Soc.* **113**: 123.
 MILLS, B. Y. (1952*a*).—*Aust. J. Sci. Res.* **A5**: 266.
 MILLS, B. Y. (1952*b*).—*Aust. J. Sci. Res.* **A5**: 456.
 MILLS, B. Y. (1952*c*).—*Nature* **170**: 1063.
 RYLE, M., SMITH, F. G., and ELSMORE, B. (1950).—*Mon. Not. R. Astr. Soc.* **110**: 508.
 SHAIN, C. A. (1954).—*Aust. J. Phys.* **7**: 150.
 SHAIN, C. A., and HIGGINS, C. A. (1954).—*Aust. J. Phys.* **7**: 130.
 STANLEY, G. J., and SLEE, O. B. (1950).—*Aust. J. Sci. Res.* **A3**: 234.

OBSERVATIONS OF THE GENERAL BACKGROUND AND DISCRETE SOURCES OF 18.3 Mc/s COSMIC NOISE

By C. A. SHAIN* and C. S. HIGGINS*

[Manuscript received September 14, 1953]

Summary

A survey of a broad strip of the sky, centred on Dec. -32° , has been made at a frequency of 18.3 Mc/s using an aerial with an overall beam width to half-power of 17° . Previous results concerning the background distribution of brightness have been confirmed and 37 discrete sources have been detected. The distribution of these sources shows some galactic concentration; it becomes homogeneous if sources within 18° of the galactic plane are excluded. From observations of source scintillations, it is concluded that some of the discrete sources have angular sizes of the order of 1° . No correlation was found between the occurrence of scintillations and published ionospheric data, but the observations are consistent with an origin of the scintillations in irregularities, of dimensions about 4 km, at a height of about 500 km.

I. INTRODUCTION

A previous paper (Shain 1951) described observations of 18.3 Mc/s cosmic noise which were made using an aerial with the direction of maximum sensitivity fixed vertically upwards in latitude 34° S. The aerial thus received cosmic noise averaged over a rather wide strip of the sky centred on Dec. -34° . It was shown, by comparison with the results of Bolton and Westfold (1950), that the observed variations in intensity as the aerial scanned this strip (which included the centre and the south pole of the Galaxy) could be explained if contours of equal intensity at 18.3 Mc/s were of the same shape as at 100 Mc/s, the absolute intensity being much higher, and if, also, the ratio of the intensity at 18.3 Mc/s to the intensity at 100 Mc/s near the galactic centre were somewhat lower than the corresponding ratio away from the centre.

Although these observations gave some indication of the way in which the intensity of cosmic noise at 18.3 Mc/s would vary over the sky, it was apparent that a detailed survey should be made with an aerial of smaller beam width, at least over the important regions near the galactic centre and at some distance from the centre. Accordingly, an aerial was constructed, to operate at 18.3 Mc/s, with a beam width of 17° between half-power points, and for which the main lobe could be swung at least 20° north and south of the zenith.

The present paper describes the observations of cosmic noise made with this aerial, these observations being mainly concentrated in the range of declination from -12° to -52° . The intensities observed were of the same order as those previously reported. With the former aerial system it was thought that the effects of one or two discrete sources of radiation could be detected; with the

* Division of Radiophysics, C.S.I.R.O., University Grounds, Sydney.

greater aerial gain and better angular resolution of the equipment described in the present paper, 37 objects have been detected whose angular size is apparently considerably less than the aerial beam width. A number of these coincide in position, within the experimental uncertainty, with discrete sources listed by Mills (1952*a*). Some of the discrete sources fluctuate in intensity although others do not, and an attempt has been made to correlate the intensity of these fluctuations with the angular sizes of the sources and with various conditions of the ionosphere.

A second paper will describe comparisons between the observational results of this paper and those obtained by other authors at frequencies near 100 Mc/s. The earlier observations showed that useful information concerning ionospheric absorption could be obtained by studying the variations in the received intensity of cosmic noise from a region of the sky under different ionospheric conditions. In the course of the present work a more detailed investigation of the effects of the ionosphere on the intensity of cosmic noise has been made, but discussion of these effects will be reported elsewhere.

II. EQUIPMENT

The equipment was situated at Hornsby, near Sydney, N.S.W. (lat. 34 °S., long. 151 °E.). The aerial site was flat but sloped downwards to the north at an angle of 2°. It was surrounded by hills at a distance of about 1 km which subtended angles of less than 10° at the aerial.

(a) *Aerial*

The aerial consisted of an array of 30 horizontal half-wave dipoles 0.2 wavelengths above ground and arranged in plan as shown in Figure 1. It should be noted that the "north-south" rows of dipoles were actually in lines directed 4.7° west of north. (This "azimuth error" has been allowed for in all the observations.) The unequal spacing between rows in the east-west direction was adopted to reduce side lobes in this direction. It was expected that, since the array was fixed and of large physical size, it would be impracticable to make a complete measurement of the aerial sensitivity pattern and that this pattern would have to be obtained by calculation, assuming the currents in the dipoles to be known. Precautions were therefore taken to ensure that the currents in each dipole were of equal amplitude and of known phase.

For each dipole a separate coaxial feeder, with a Pawsey stub as balance-unbalance transformer, ran to a hut in the centre of the array. As the lengths of feeder were large, about 100 ft, the attenuations in the cables feeding each dipole were equalized as far as possible by making all the feeders from the north-south rows 2, 3, 4, and 5 (see Fig. 1) of the array the same length (five half wavelengths) and these from rows 1 and 6 one wavelength longer.

At the centre hut the feeders were first connected in threes, e.g. A1, A2, A3 ; B1, B2, B3 ; etc., and the impedance of each combination was matched, using a T-network, to the characteristic impedance of the cable used for the feeders. The outputs from the five matching networks in the western half of the aerial were then combined and the impedance of the combination again matched to

the feeder characteristic impedance (see Fig. 2 (a)). Similarly, the feeders from the eastern half of the array were combined and two cables, one from the eastern half and one from the western half of the array, were taken to a hut outside the array which contained the receivers and recording equipment.

In order to direct the aerial beam to different declinations appropriate extra lengths of cable were inserted at the points indicated by small circles in Figure 2 (a).

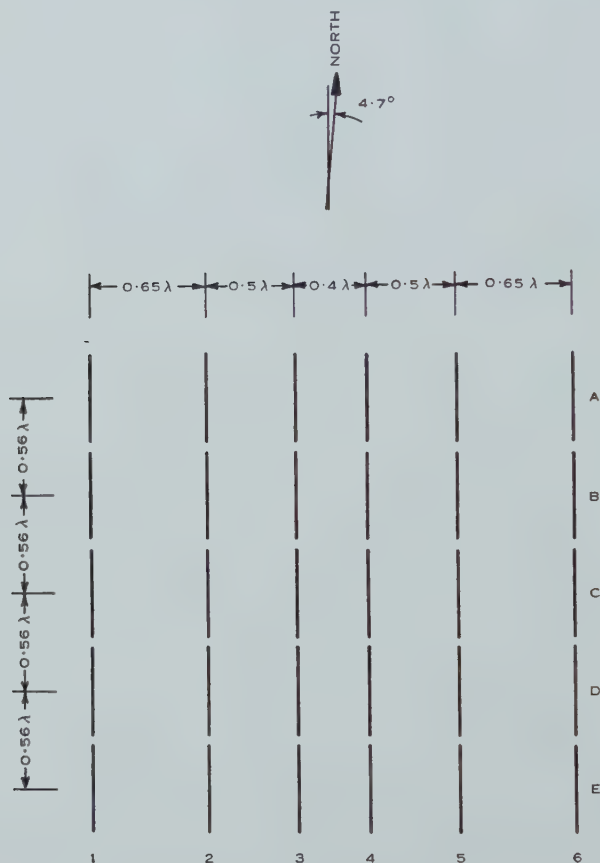


Fig. 1.—Plan of aerial array. The rows of dipoles are numbered and lettered for reference to individual dipoles.

In the recording hut the two feeders were connected to a "feeder bridge" (Westcott 1948) as shown in Figure 2 (b), and two receivers were connected to the other corners of the bridge. The feature of this bridge is that the two receivers are quite independent of each other in operation, and one receiver (receiver 1 in Fig. 2 (b)) uses the aerial with the eastern and the western halves in phase while the other (receiver 2) has the two halves out of phase. The use of the two receivers simultaneously was helpful in recognizing some interfering signals and also permitted accurate direction-finding observations in some

circumstances. The latter facility was used particularly in observations of solar noise to be described in another paper.

In setting up, each dipole was matched to its feeder with all other dipoles open-circuited. All feeders were then joined together and the whole array excited in phase. The impedance of each dipole was measured by observation of the standing waves in the appropriate feeder. This was done by inserting a special length of cable one wavelength long and comparing the relative amplitudes of the voltage in the line at three points spaced one-eighth of a wavelength apart. (A similar system has been described recently by Sutcliffe (1953).) It was found that mutual impedances had changed the dipole impedances slightly,

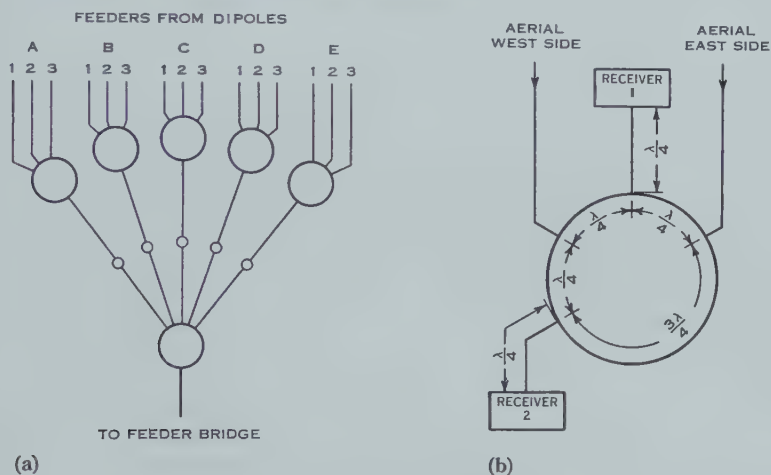


Fig. 2 (a).—Matching arrangements for feeders from the western half of the array. The large circles indicate the matching networks and the small circles the points where the extra lengths of cable were inserted for beam-swinging. Fig. 2 (b).—Arrangement of feeder bridge. All the feeders were of the same type of coaxial line, having a nominal impedance of $70\ \Omega$, except the two quarter-wave sections at the receiver inputs which had a nominal impedance of $50\ \Omega$.

but that all dipole impedances were very nearly the same. Finally, the feeders were connected in working order and the junction matching networks adjusted in turn. The aerial impedance measurements were repeated a number of times during the period of the observations.

An overall check of the equality of the feeders in the eastern and western halves of the array was obtained using a small oscillator attached to a balloon. Owing to the difficulty of controlling the balloon, this system could not be used to determine the detailed aerial polar diagram, but it was shown that the characteristics of the aerial which could be checked were as expected. A rough check of the positions of some side lobes and of the minima of the aerial diagram was also obtained during recordings of several intense solar disturbances.

Typical calculated aerial diagrams in the north-south plane are shown in Figure 3 (a) and the two diagrams for the east-west plane in Figure 3 (b). It

should be noted that, because of the slope of the aerial site in the north-south plane, the aerial beam was directed to Dec. -32° rather than -34° , when no extra lengths of cable were inserted in the feeder system for beam swinging.

It was calculated that with the aerial beam in its normal position the proportion of power in the main lobe was 89 per cent., this proportion decreasing to 84 per cent. when the beam was swung through 20° .

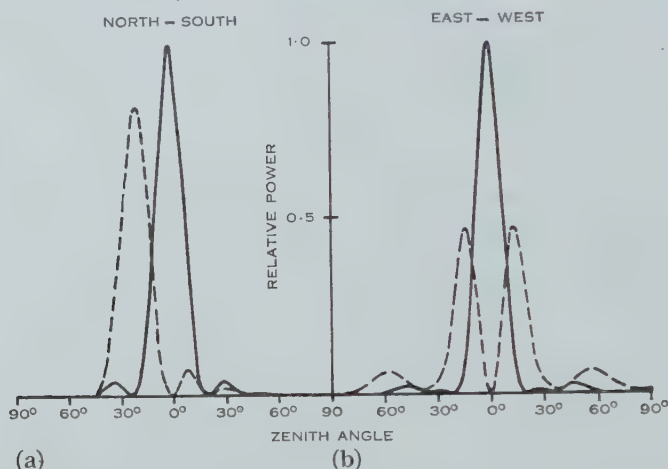


Fig. 3.—Aerial sensitivity. (a) North-south direction. Patterns are shown for the aerial directed to declinations -32° (full line) and -12° (dashed line). (b) East-west direction. The pattern used by receiver 1 (see Fig. 2 (b)) is indicated by the full line and that for receiver 2 by the dashed line.

(b) Receivers and Calibration

The two receivers were standard communication-type receivers with bandwidths of the order of 1 kc/s. As the intensity of the cosmic noise was always very high compared with the intensity of the noise generated in the receivers, no great care was taken to improve the receiver noise factors. Recording microammeters measured the detector current in each receiver.

Calibration of the received noise intensity was provided by a noise generator using a CV172 noise diode, the output impedance of the noise generator being adjusted to be equal to the impedance presented to the receiver by the aerial feeders.

(c) Losses in the Aerial System

The measured values of the noise power at the receivers were corrected for the power lost in the ground and in the aerial feeder system.

The power loss in the ground was calculated using appropriate values of conductivity and dielectric constant. The calculated loss, which does not depend critically on the values assumed for these constants, was 1.6 db.

The power loss in the feeder system was measured by observing the standing wave pattern in the feeder near the receivers when the dipole ends of the feeders

were short-circuited. Also a direct measurement was made of the losses between the first matching networks and the receivers, using the noise generator. Allowing for the loss in the extra cable between the matching networks and the dipoles, this gave another estimate of the feeder losses which confirmed the first. The value adopted was 2.6 db.

Thus the total losses in the aerial system were estimated to be 4.2 db with an uncertainty of about ± 0.3 db (that is, about ± 7 per cent.).

III. OBSERVATIONS

(a) *Observational Procedure*

The equipment was run, almost continuously, during the period June 1950 to June 1951. The observations described in the present paper are largely based on a number of accurate records for which the receivers were run at a comparatively high gain, with part of the detector current backed off. While these records were being taken the receivers were monitored continuously and calibration signals were put on the records at intervals of less than 1 hr. These records were taken in several series at intervals of a few months and, during each series, records were taken on at least two nights with the aerial directed to each of the declinations -12° , -22° , -32° , -42° , and -52° . After each of these series of observations, readings of noise power were taken from the records at intervals of about 6 min and plotted as equivalent aerial temperatures against sidereal time.

At other times the receivers were left running with somewhat lower gain. At infrequent intervals the receivers were manually disconnected in turn from the feeder bridge and connected to the noise generator while a series of known noise intensities was put on the record. Between calibrations the receivers were usually unattended. Some interference was experienced from distant radio stations transmitting on the frequency to which the receiver was tuned (this was always within 25 kc/s of 18.3 Mc/s), and sometimes the record was lost during thunderstorms occurring within a radius of several hundred miles of the receiving site. However, these interfering signals could be recognized on the record and the recording time lost from these causes was small, so that records were obtained for practically every day during the year's observations. Having the two receivers tuned to slightly different frequencies often helped in deciding whether a peculiarity in the record was due to natural causes, for example solar noise, or to station interference.

(b) *Ionospheric Attenuation*

From experience in the earlier observations, it was expected that ionospheric absorption would be serious when the critical frequency of the F_2 region of the ionosphere was higher than 9 Mc/s, and with the improved recording technique it was thought that the highest critical frequency that could be tolerated would be somewhat less than 9 Mc/s. Absorption in the D region could be neglected since only records taken at night were used.

When records taken on different nights were compared, it was found that the equivalent aerial temperatures recorded for the same sidereal time were the

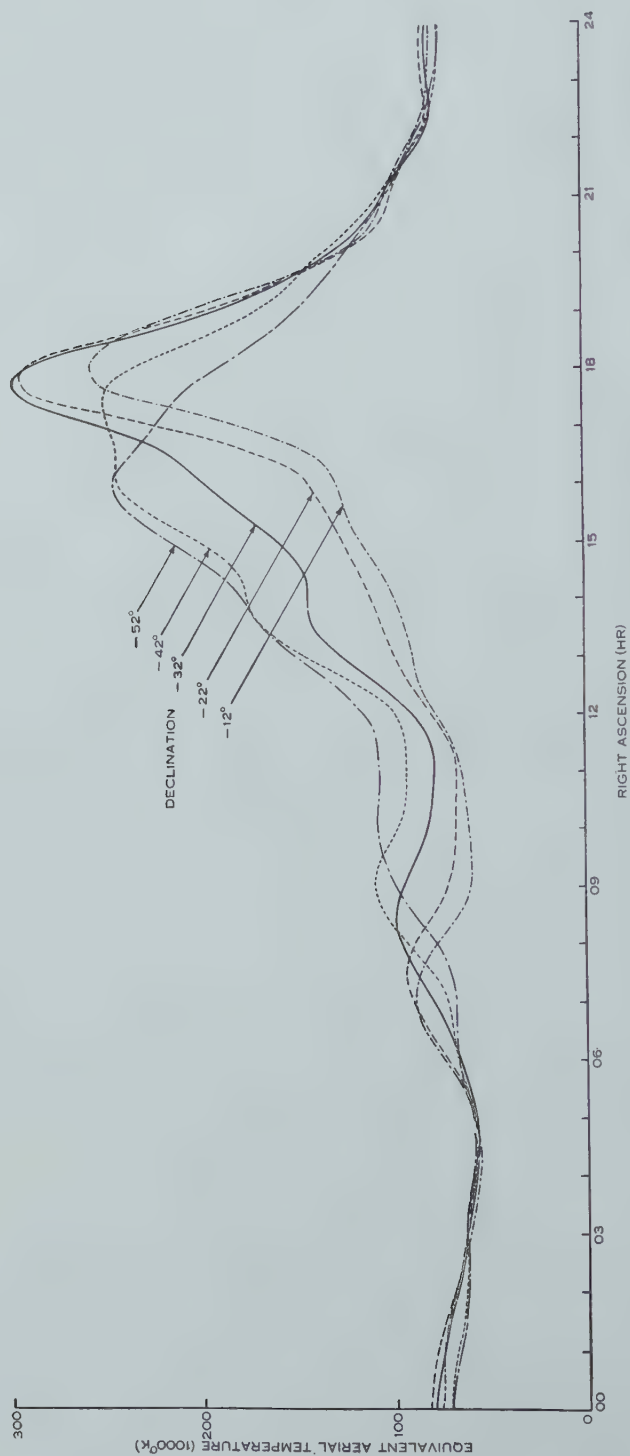


Fig. 4.—Observed equivalent aerial temperatures as functions of Right Ascension for the aerial beam directed to the five declinations indicated on the curves.

same within the experimental uncertainty so long as the critical frequency of the F_2 region was less than 5 Mc/s. Accordingly, only records taken under this condition were used in the main analysis and most of the zone of the sky observed could be covered with the more accurate records. The gaps in these records, during which the F_2 region critical frequency was greater than 5 Mc/s, were filled in using all the lower gain records for which the ionospheric conditions were suitable.

(c) Results

The results of these observations, as equivalent aerial temperatures plotted against Right Ascension for the five declinations, are shown in Figure 4. The estimated probable error in relative values of equivalent aerial temperature is about ± 2 per cent. and the probable error in the absolute values about ± 10 per cent. The results have been replotted in Figure 5 as contours of equal equivalent aerial temperature, using galactic coordinates.*

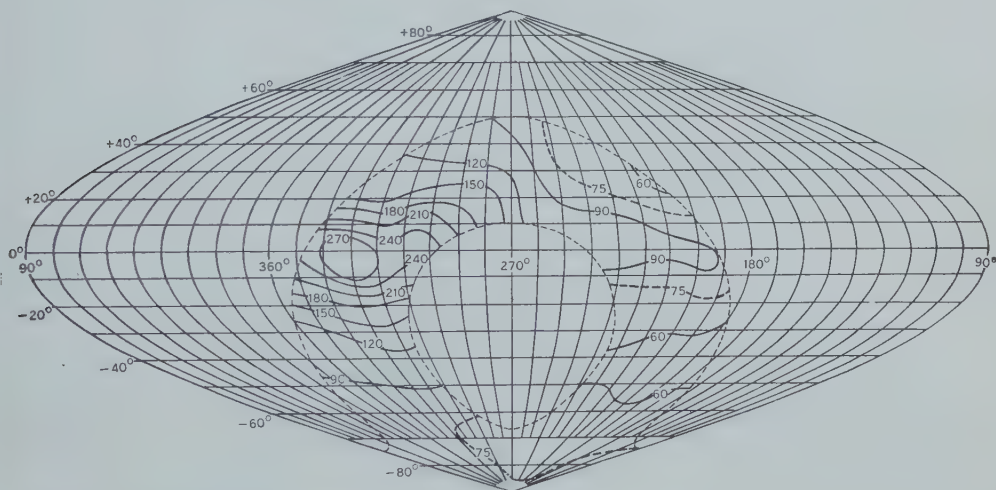


Fig. 5.—Contours of equivalent aerial temperature (in units of 1000 °K), plotted in galactic coordinates.

IV. ANALYSIS

(a) Brightness Distribution

Comparison of Figures 4 and 5 with corresponding figures in the earlier paper (Shain 1951) shows that the observed equivalent temperatures are of the same order but somewhat higher. The ratio of maximum to minimum for the present observations is greater than the ratio observed previously, owing to the narrower aerial beam used, but also the absolute intensities have been raised slightly, owing mainly to a better estimate of the aerial system losses.

* The values of Right Ascension and declination in Figure 4 are for the period of the observations. For the conversion to galactic coordinates (Fig. 5) the adopted position of the north galactic pole was R.A. 12 hr 40 min, Dec. $+28^\circ$ (1900).

A detailed comparison between the present observational results and those of Bolton and Westfold (1950) will be made in another paper. However, a rough comparison of Figure 5 with the contours of observed equivalent temperatures at 100 Mc/s given by Bolton and Westfold shows that at both frequencies there is the same general trend towards high intensities near the galactic centre and that the minimum intensities are in approximately the same position. Local differences in the shapes of the contours, for example in the regions near $l=280^\circ$, $b=+20^\circ$ and $l=305^\circ$, $b=+5^\circ$, are probably due to the effects of discrete sources.

Stanley and Slee (1950) have shown that the flux density from a number of discrete sources varies approximately in proportion to the wavelength, while the background brightness temperature varies approximately as the wavelength to the power 2.5, that is, the background brightness (flux per unit solid angle) varies as the wavelength to the power 0.5. Therefore, provided aerials of the same beam width are used, and this is approximately true when comparing the present observations with those of Bolton and Westfold, the relative increase in the received power due to discrete sources compared with the background power should vary approximately as the square root of the wavelength. Although it will be shown in the subsequent paper that, in the range of frequencies between 18.3 Mc/s and 100 Mc/s, the variation of source flux density with wavelength is somewhat different from that deduced by Stanley and Slee, it is in fact more rapid than the variation of background brightness with wavelength. This would result in a distortion of the contour lines as observed.

(b) *The Discrete Sources—Detection*

The irregularity in the contours of Figure 5 in the region near $l=280^\circ$, $b=+20^\circ$ corresponds to the increase shown near R.A. 13 hr 30 min on the curve of Figure 4 for Dec. -42° and this agrees with the position of the discrete source Centaurus-A (Stanley and Slee 1950). A number of other similar increases could be detected easily. Since the objects giving rise to the localized increases in intensity must subtend an angle somewhat less than the beam width of the aerial and some at least correspond in position with discrete sources observed at higher frequencies, they have been called discrete sources; it should be remembered that they could have an angular size of several degrees. However, associated with some of these increases were short duration (order of 1 min) fluctuations in the received intensity and, as discussed later, this observation suggested that some at least of the discrete sources would subtend angles of the order of 1° or less.*

A careful check of the observational results was made for effects of discrete sources. The procedure, in the case of three of the weaker sources, is illustrated

* Evidence that the increase corresponding to the source Centaurus-A is due to a source of radiation having an angular size of the order of 1° was obtained during a short series of observations using the "sea interferometer" technique. Under suitable ionospheric conditions an interference pattern was obtained using an array of 12 half-wave dipoles erected near the top of a hill about 600 ft high and close to the sea. However, owing to the uncertain amount of ionospheric refraction and attenuation, these observations could not be used to improve the estimates of the position and intensity of the source obtained by other methods.

in Figure 6. In this figure the curves on which part of Figure 4 is based have been redrawn on a larger scale and with the curves for three declinations displaced vertically for clarity. The dotted sections of the curves correspond to the periods of the records during which fluctuations were observed. For each declination, records from observations on at least two separate days were averaged; the daily records were so similar that they could not be separated in the figure. Certain small "bumps" on the curves repeated themselves each

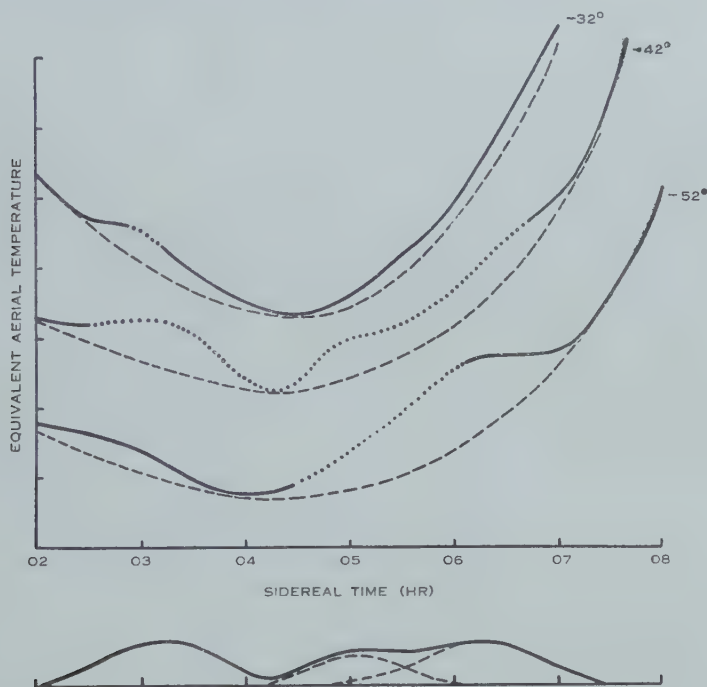


Fig. 6.—The detection of discrete sources. In the upper part of the figure are shown experimental curves (full and dotted lines) of equivalent aerial temperature versus sidereal time for declinations -32° , -42° , and -52° . The curves are displaced vertically for clarity (cf. Fig. 4). The dotted part of the curve indicates the periods during which fluctuations were observed and the dashed lines indicate the estimated variation of the general background temperature with sidereal time. The lower part of the figure shows the estimated discrete source contributions for Dec. -42° . The passage of three sources through the aerial beam is indicated (sources S03-4, S05-4, and S06-5 in Table 1).

day and there is apparently a progressive change with declination in the character of the records. These bumps, and the appearance of fluctuations, indicate the presence of several discrete sources and a rough indication of their position and intensity can be obtained directly from the figure. To obtain a better estimate of the positions and intensities, the estimated trend of the general background radiation was drawn for each declination (dashed lines in Fig. 6) and the differences between the observed temperatures and the estimated background

TABLE 1
LIST OF DISCRETE SOURCES OBSERVED AT 18.3 mc/s

Number	R.A. (hr min)	Dec. (deg)	l (deg)	b (deg)	Flux Density (S) ($10^{-25} \text{W m}^{-2} (\text{c/s})^{-1}$)	Level ($L = 25 + \log_{10} S$)	Notes
S							
00—1	00 23	—19	65	—81	1500	3.2	
01—2	01 34	—20	150	—75	1100	3.0	
03—4	03 24	—41	211	—54	350	2.5	
04—2	04 22	—21	185	—39	700	2.8	
05—4	05 12	—45	217	—35	350	2.5	
06—1	06 05	—18	192	—16	1000	3.0	
06—3	06 22	—33	209	—18	300	2.5	
06—5	06 21	—51	227	—24	850	2.9	
07—1	07 10	—18	199	— 2	1200	3.1	
08—4	08 20	—43	228	— 3	700	2.8	
09—1	09 23	—12	214	+28	400	2.6	
09—3	09 20	—34	230	+12	700	2.8	
09—5	09 55	—53	246	+ 2	800	2.9	
10+1	10 45	+16	199	+61	1300	3.1	
10—1	10 20	—10	224	+39	800	2.9	
12+1	12 30	+11	262	+73	1100	3.0	
12—0	12 53	— 1	277	+61	4100	3.6	(1)
12—1	12 24	—14	263	+47	1500	3.2	
13—2	13 20	—22	284	+40	2000	3.3	
13—3	13 44	—33	285	+27	650	2.8	
13—4	13 25	—43	279	+18	5300	3.7	(2)
13—5	13 00	—58	273	+ 5	3300	3.5	(3)
14—4	14 45	—42	294	+15	5300	3.7	
15—1	15 21	—16	317	+31	1000	3.0	
15—6	15 35	—61	290	— 5	6300	3.8	(3)
16—3	16 02	—39	308	+ 8	3800	3.6	
18—2	18 20	—28	332	— 9	5700	3.8	
19+0	19 30	+ 2	8	—10	1100	3.0	
19—2	19 40	—20	349	—23	650	2.8	
20+x	20 02	+ x			(66000)	(4.8)	(4)
20—4	20 06	—47	320	—34	350	2.5	
21—1	21 20	—15	5	—42	1400	3.1	
21—2	21 50	—25	354	—52	500	2.7	
21—3	21 05	—30	344	—44	800	2.9	
21—5	21 56	—52	311	—51	250	2.4	
23—2	23 20	—25	5	—72	1550	3.2	
23—5	23 30	—51	295	—64	250	2.4	

(1) Intensity uncertain owing to confusion with S12—1.

(2) Centaurus-A. Probable error in position ± 2 min in Right Ascension, $\pm 1^\circ$ in declination.

(3) Uncertainty in declination. Intensity depends on adopted position.

(4) A small increase observed with the aerial beam in furthest north position. No means of obtaining declination with any accuracy but consistent with suggestion that the increase is due to the source Cygnus-A. Intensity estimated assuming Dec. $+40^\circ$.

were plotted against sidereal time as in the lower part of Figure 6. This curve represents the variations in equivalent aerial temperature at Dec. -42° due to the passage through the aerial beam of the sources only. Knowing the relative aerial sensitivities, the relative amplitudes of the estimated source contributions at the different declinations provided a check on the consistency of the estimates of the background trend. In some cases, as on the right side of Figure 6, the source bump was of much longer duration than expected from the aerial beam width and is clearly due to the effect of two neighbouring sources.

A similar procedure was followed for all the records. Sometimes, when the rate of change of the background temperatures with Right Ascension was high, clearer indications of sources were obtained by plotting equivalent aerial temperatures against declination or against galactic latitude. In all, 31 sources could be detected in the zone bounded by declinations -4 and -60° (corresponding to the lines swept out by the half-power points of the aerial beam during the main survey). In addition, records were available for two other positions of the aerial beam and, although the aerial gain was lower in these positions, a further six sources could be detected.

In general, for all the sources, positions could be assigned only roughly, with an estimated probable error of up to 5° for the weaker sources, although the uncertainty was somewhat lower for the stronger sources.* Similarly, the estimated probable errors in the intensities vary from about ± 20 per cent. for the strongest to a factor of 2 for a few of the weakest sources.

A list of the sources detected is given in Table 1. The sources have been numbered according to the system used by Mills (1952*a*). In this system the hour of Right Ascension is followed by the sign and tens of degrees of the declination, so that the designation of a source gives some indication of its position. To avoid confusion with the sources in Mills's list, the numbers of the sources in Table 1 will be prefixed by "S" (e.g. S05-4).

Also in accordance with the suggestion of Mills, the "level", L , of a source is defined by $L = \log_{10}(S \times 10^{25})$, where S is the flux density of the source (total for two planes of polarization) in units $(\text{W m}^{-2} (\text{c/s})^{-1})$.

A map of the sources, using galactic coordinates, is given in Figure 7. Each source is represented by a symbol which gives an indication of its level. Also shown in Figure 7 are open circles showing the positions of all the sources detected by Mills within the regions of the sky of interest. It is seen that a number of the sources listed in Table 1 have positions close to sources listed by Mills and, taking into account the uncertainties in the positions of the sources, dashed lines have been drawn round the symbols which it is thought probably represent observations of the same object at the two frequencies.

* The position of the source Centaurus-A was obtained more accurately, with a probable error of about ± 1 min in Right Ascension and $\pm 1^\circ$ in declination, by comparing the amplitudes of fluctuations in intensity as observed simultaneously with the two receivers, using different aerial diagrams as shown in Figure 3 (*b*).

(c) *Distribution of the Discrete Sources*

The distribution of the sources detected in the range of declinations from -4 to -60° has been analysed by a method similar to that used by Mills. The logarithm of the number of sources with level L or greater is considered as a function of L . If there is a random distribution of a large number of sources, this function should be a straight line having a slope -1.5 . Figure 8 shows the ogive for the sources observed at 18.3 Mc/s. The points corresponding to levels 2.6 or less are probably not very reliable since the numbers of sources observed having these levels will be less than the true numbers owing to confusion with sources of greater intensity. It is seen that the points are fitted by a line

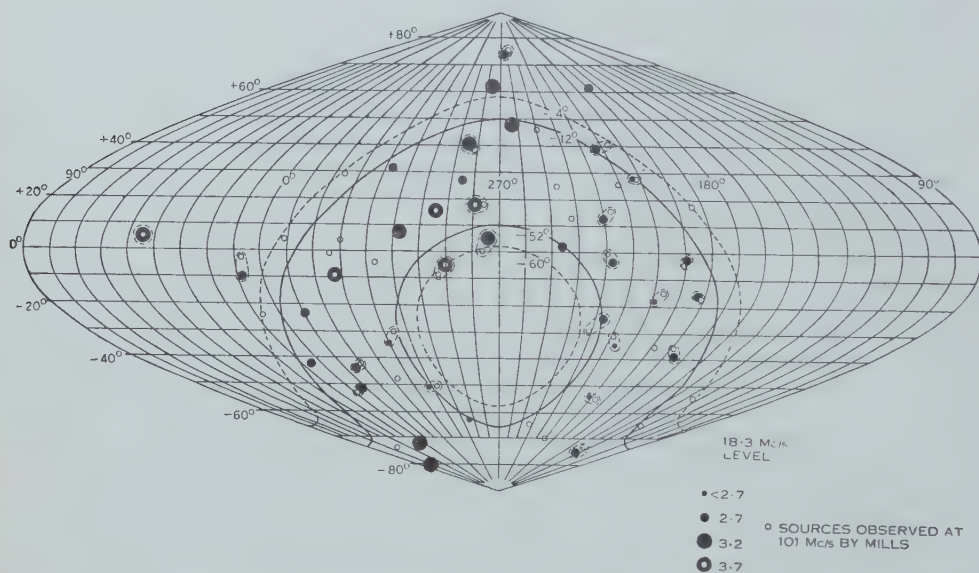


Fig. 7.—Map of the discrete sources observed at 18.3 Mc/s (galactic coordinates). Intensities are indicated by different symbols as shown. The full and dotted curves show the limits of the aerial maximum and half-power points in the main survey. Open circles give the positions of sources observed by Mills; dashed lines have been drawn round the symbols which probably indicate observations of the same source.

having a slope of about -1.1 rather than -1.5 . This result is similar to that obtained by Mills. Mills then considered separate ogives for the sources within $\pm 12^\circ$ of the galactic plane and for the remainder. He found that the sources near the galactic plane gave an ogive slope of -0.75 while the remaining sources gave a slope of nearly -1.5 , a result consistent with the hypothesis that the latter sources comprised a homogeneous distribution.

Although, when analyzed in the same manner, the 18.3 Mc/s sources gave a similar result, it was thought that, with the smaller number of sources available for discussion, a different method of presentation, involving the largest possible number of sources at each step, should be adopted. First, the least squares solution for the slope of the number-level ogive was calculated, omitting sources with level 2.6 or less. Then the three sources in the band around the equator

from latitudes -3 to $+3^\circ$ were removed and the slope of the ogive for the remaining sources was calculated. This procedure was repeated as the band around the equator containing the discarded sources was widened in small steps until the number of sources left became small. (For latitude limits 0, 18, and 27° the numbers of sources used were respectively 24, 14, and 11.) The results of these calculations are shown in Figure 9 where the points represent the slope of the ogive neglecting the sources in a band around the equator extending to the appropriate latitude limit. It is seen that the magnitudes of the slopes are about 1.3 or less until the ogive is restricted to sources outside a range of 18° from the galactic plane and then increases sharply to at least 1.5. This suggests that the sources outside that range of latitude are distributed homogeneously

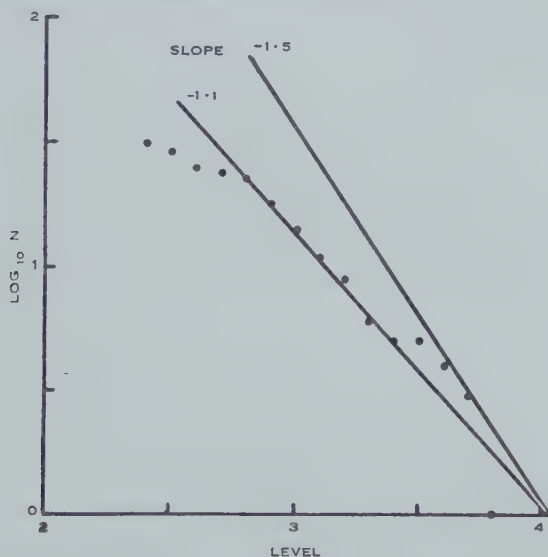


Fig. 8.—Number-level ogive for all sources with declinations between -4 and -60° . N is the number of sources with level L or greater.

whereas the sources within $\pm 18^\circ$ of the galactic plane include some other class of source. This agrees with Mills's conclusions. Similar calculations to the above were made for the sources in Mills's list (omitting sources with levels less than 1.0) and these gave the slopes indicated by the crosses in Figure 9. The 18.3 Mc/s and 101 Mc/s results are in very good agreement.

A check was made to see whether the density of sources was greater in the northern galactic hemisphere than in the southern hemisphere. It was found that there was no significant difference, but the comparatively small area of the northern hemisphere covered in the survey did not allow a very sensitive test to be made.

(d) Source Scintillations

(i) *Observations.*—Soon after recording began it was noticed that there were often irregular fluctuations in intensity at times when it appeared that a discrete source was passing through the aerial beam, although there were some

sources which were rarely, if ever, seen to fluctuate in intensity. It was thought that the presence or absence of scintillations might be a property of a source. However, Ryle and Hewish (1950) have found a diurnal variation in the amplitude of these "scintillations" of sources observed at high angles while Bolton, Slee, and Stanley (1953) found both annual and diurnal variations for their low angle observations and it was therefore necessary to check whether the lack of scintillations for some sources was not merely an effect of the times of observation. Unfortunately there are no sources for which observations extend over more than a few months (corresponding to a few hours' variation in local time of observation), but a number of sources were chosen for study of their scintillations so that there were observations covering all hours of the day and a large part of the year.

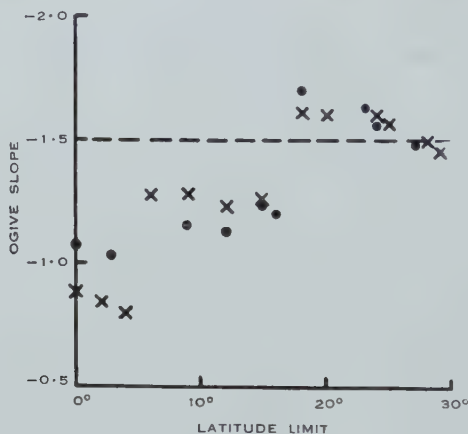


Fig. 9.—Slope of the number-level ogive for sources outside a band around the galactic equator having the width specified by the appropriate latitude limit.

• 18.3 Mc/s: sources between -4 and -60° declination with levels 2.7 or greater.

× 101 Mc/s: all sources with level 1.0 or greater.

The average amplitude of the scintillations for a record of a particular source was estimated in terms of the amplitude of the inherent noise fluctuations of the trace. For all the observations the bandwidth of the receiver was kept constant and similar recorders were used. Then, although the records were made with various settings of the receiver gain, the width of the trace on the record corresponded to a constant fraction of the total noise intensity at the time of observation. The estimates of the amplitude of the scintillations could then be converted to the units in which the intensity of the source was expressed. A scintillation index for observations of a source during a particular month was calculated as one-half the monthly average of the estimated daily averages of the amplitude of the scintillations divided by the intensity of the source. This index should then be comparable with that used by Ryle and Hewish. Although the method of estimation is rather rough, it is sufficiently reliable for comparison between different sources and for correlations of scintillations with ionospheric phenomena.

(ii) *Results*.—Scintillation indices were calculated, for all months for which observations were available, for 12 sources. In Figure 10 (a) these indices are plotted against the month of observation and in Figure 10 (b) against time of

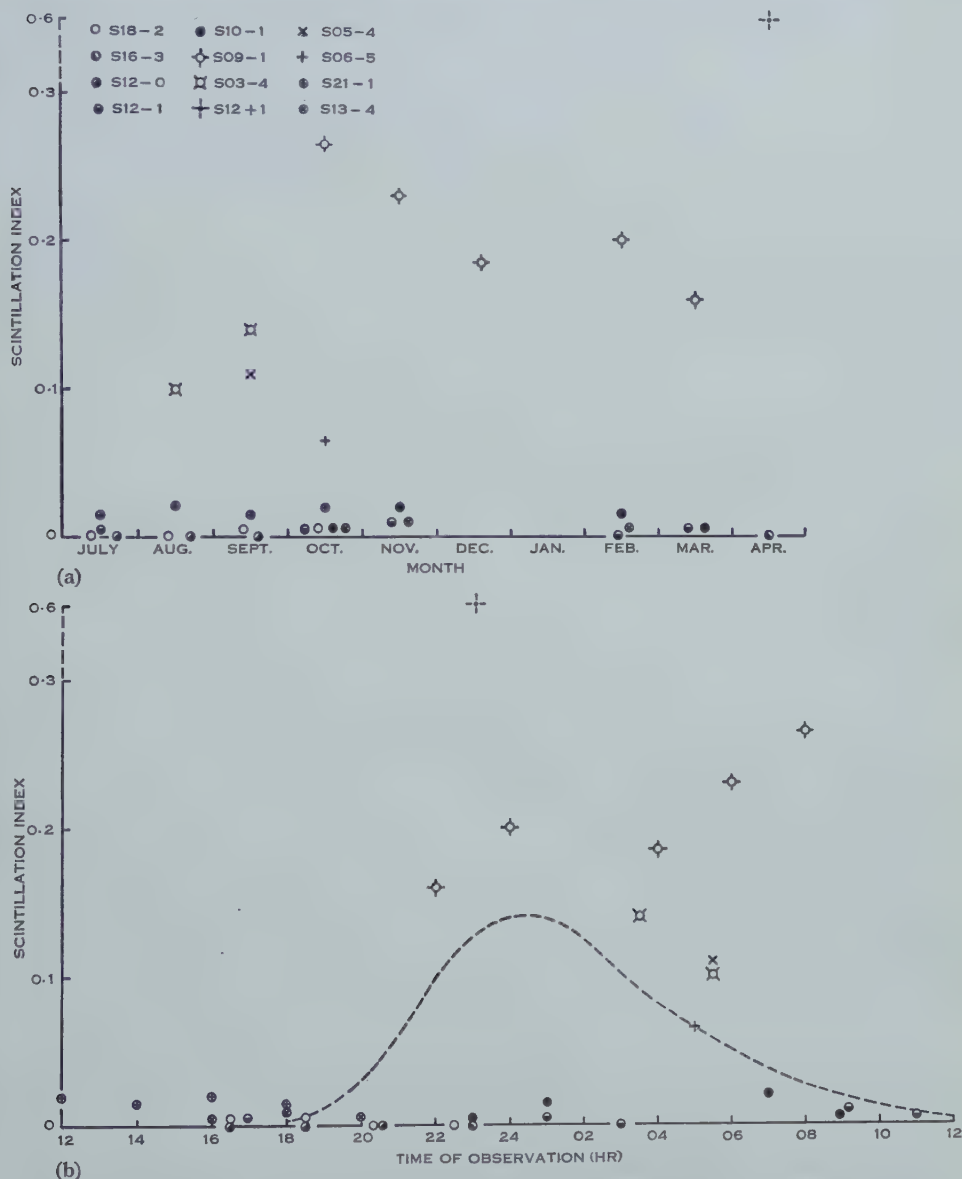


Fig. 10.—Scintillation indices for 12 sources (each represented by a symbol as shown) plotted against (a) month of observation, (b) time of day. The dashed curve in (b) shows the diurnal variation of scintillation index observed by Ryle and Hewish (1950).

day. The dashed curve in Figure 10 (b) represents the diurnal variation of scintillation index reported by Ryle and Hewish from their observations of four northern sources. Both figures show a considerable scatter in the values of

scintillation index. Although high values of scintillation index have not been observed during the winter nor during the midday and afternoon hours, this does not imply marked diurnal and seasonal variations since the sources with high scintillation indices have not been observed at these times. In fact no significant trends can be deduced from these data, but Figures 10 (a) and 10 (b) together bring out the fact that there are wide differences in the scintillation indices for different sources, even when observed at approximately the same time of day in the same season. This can be seen in the differences between the sources S09-1, S10-1, and S12-1, but the most striking illustration is the difference between the sources S12+1 and S12-0.

These two sources appear on some records as one clear increase (bump) somewhat lengthened with respect to the aerial beam width (see Fig. 11). On

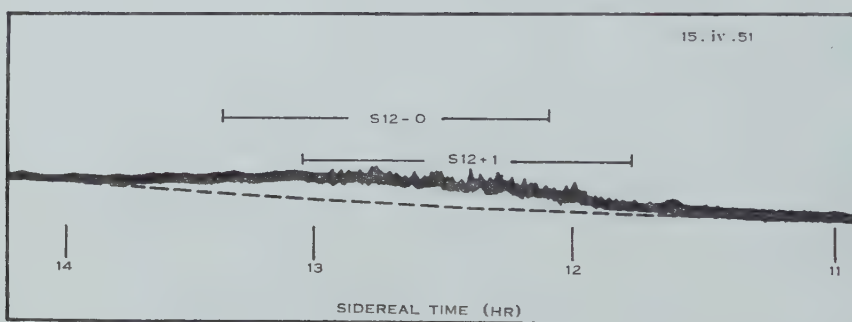


Fig. 11.—A record showing the “bump” as the sources S12+1 and S12-0 pass through the aerial beam. The source S12+1 which transits earlier shows marked scintillations, whereas there are few, if any, scintillations on the later portion of the bump, due to the source S12-0. The lines above the trace show the half-power points for the two sources; the dashed line indicates the background radiation.

this bump fluctuations in intensity were often observed for a time which would correspond to the passage of a single source through the aerial beam, but the maximum of the fluctuations was clearly displaced from the maximum of the smoothed bump. Thus, of the two sources observed under practically the same conditions, one shows considerable scintillations whereas the other shows none.*

The suggested explanation of this difference is to be found in the different angular sizes of the sources. Previous work (e.g. Little and Maxwell 1951) has suggested that the scintillations are due to diffraction by irregularities in the *F* region of the ionosphere, the irregularities having such a size as to subtend an angle of about $0.5-1.0^\circ$ at the Earth's surface. Then, for constant observing conditions, a source will have a high or low scintillation index depending on whether its angular size is much less or greater than about 0.5° . The sources considered in Figure 10 may be roughly divided, according to scintillation indices, into the classes shown in Table 2, corresponding to scintillation indices

* The source S12+1, together with S09-1, is also distinguished by the more rapid rate of scintillation (up to several peaks per minute) in comparison with the other sources (one peak every few minutes).

0.01 or less, 0.02, 0.1, and greater than 0.2. Then, if the critical angular size for the sources to show marked scintillations is 0.5° , the classes in Table 2 would correspond to angular sizes of sources 1° or greater, 0.5 – 1.0° , about 0.5° , and less than 0.5° respectively. Although there may be some doubt as to the absolute scale of sizes, the table should give a good indication of relative angular sizes.

The source S12+1 corresponds to the source observed by Mills (1952*a*) and designated 12+1 (Virgo-A). This source has been shown (Mills 1952*b*) to have an equivalent angular size of about 0.1° at 101 Mc/s, and in recent observations Mills (personal communication), using a new technique, has detected a source close to the position of S12-0 which has an angular size greater than 1° . Assuming the source sizes at 101 Mc/s are comparable to those at 18.3 Mc/s, Mills's observations agree with the values deduced from Table 2 and confirm the suggested explanation for the differences in scintillation indices.

TABLE 2

OBSERVED SCINTILLATION INDEX (A ROUGH INDICATION OF ANGULAR SIZE)

≤ 0.01	0.02	0.1	> 0.2
S12-0	S10-1	S03-4	S09-1
S12-1	S13-4	S05-4	S12+1
S16-3	S21-1	S06-5	
S18-2			

(iii) *Correlation between Scintillation Amplitude and Ionospheric Data.*—The daily amplitude of the scintillations of the sources S09-1 and S13-4 were compared with fE_s (the highest frequency on which "sporadic E" echoes are observed) and with the occurrence of "spread F" echoes, from the records of the Canberra station of the Ionospheric Prediction Service of the Commonwealth Department of the Interior.

For correlation with fE_s , the scintillation amplitudes were divided into several ranges and the average was found of the values of fE_s recorded during the observations of scintillations in a certain range. No correlation was obtained between fE_s and amplitude of scintillations. Also, using a similar method, no correlation was found with magnetic K -index.

During the period of interest, there were very few occasions on which spread F echoes were reported and on these occasions the amplitude of the scintillations was neither unusually high nor low. Nevertheless it is possible that the occurrence of scintillations is connected with the occurrence of spread F echoes as the amount of "spread" in critical frequency of the irregularities required to produce scintillations at 18.3 Mc/s is probably so small (see, e.g. Little 1951) that it would not be noted in routine ionospheric recording.

(iv) *The Height of the Irregularities causing the Scintillations.*—Provided the lateral dimensions of the irregularities causing the scintillations are greater than the dimensions of the first Fresnel zone at the height concerned, for radiation

of the lowest frequency considered, the scintillation index would be expected to vary as the square of the wavelength. Such a variation was found by Hewish (1952) and also by Bolton, Slee, and Stanley (1953) for frequencies near 100 Mc/s. However, it will be seen from Figure 10 (*b*) that the scintillation index for the smallest (angular size) sources observed at 18.3 Mc/s is not much more than the index for sources observed at 81 Mc/s by Ryle and Hewish (1950), indicating a very slow variation with wavelength. The possible explanation of this is that when the frequency is reduced to 18.3 Mc/s, the size of the irregularities, assumed the same at all frequencies, is of the order of magnitude of the first Fresnel zone.

Several authors (e.g. Little and Maxwell 1951) have shown that the lateral dimensions of the irregularities are about 4 km. This would be the diameter of the first Fresnel zone for 18.3 Mc/s at a height of about 500 km, which is consistent with the height of the irregularities deduced by Hewish (1952). On the basis of the arguments of this Section, such a height would be an upper limit. A lower limit cannot be set definitely. For example, if the irregularities causing scintillations were in the *E* region (at a height of about 120 km), a size of 4 km would be only about twice the diameter of the first Fresnel zone, so that this height could not be excluded.

Of course these arguments break down if the irregularities causing the scintillations at higher frequencies are different from those affecting the 18.3 Mc/s radiation.

V. CONCLUSION

Although, for different frequencies there are important differences both in absolute intensities and in the distributions of brightness over the sky, these observations have confirmed earlier conclusions that the variation with frequency of the characteristics of cosmic noise is smooth, so that useful comparisons can be made between observations at comparatively widely-spaced frequencies. The observations described in the present paper should be useful for such detailed comparisons. In particular, the aerial diagram of the array used for this work bears a close similarity to that of the aerial used by Bolton and Westfold (1950) and, with some minor corrections, these two surveys can be compared directly. Such a comparison will be considered in a subsequent paper. That paper will also discuss the information that can be obtained from a comparison of the flux densities of the discrete sources given in this paper and as measured by Mills (1952*a*).

These observations were made using the pencil-beam technique, even though the resolution was not very high. They have confirmed the suggestion that a fairly large number of the discrete sources subtends considerable angles, certainly much greater than the angles subtended by visual stars, and they have also confirmed the conclusions of Mills concerning the distribution of the discrete sources over the sky. The observations of scintillation phenomena have been shown to be consistent with the deductions of previous workers from observations of the scintillations at higher frequencies.

VI. REFERENCES

- BOLTON, J. G., SLEE, O. B., and STANLEY, G. J. (1953).—*Aust. J. Phys.* **6**: 434-51.
BOLTON, J. G., and WESTFOLD, K. C. (1950).—*Aust. J. Sci. Res. A* **3**: 19-33.
HEWISH, A. (1952).—*Proc. Roy. Soc. A* **214**: 494-514.
LITTLE, C. G. (1951).—*Mon. Not. R. Astr. Soc.* **111**: 289.
LITTLE, C. G., and MAXWELL, A. (1951).—*Phil. Mag.* **42**: 267-78.
MILLS, B. Y. (1952a).—*Aust. J. Sci. Res. A* **5**: 266-87.
MILLS, B. Y. (1952b).—*Nature* **170**: 1063.
RYLE, M., and HEWISH, A. (1950).—*Mon. Not. R. Astr. Soc.* **110**: 381.
SHAIN, C. A. (1951).—*Aust. J. Sci. Res. A* **4**: 258-67.
STANLEY, G. J., and SLEE, O. B. (1950).—*Aust. J. Sci. Res. A* **3**: 234-50.
SUTCLIFFE, H. (1953).—*Wireless Engr.* **30**: 180-1.
WESTCOTT, C. H. (1948).—*Wireless Engr.* **25**: 215-20.

A COMPARISON OF THE INTENSITIES OF COSMIC NOISE OBSERVED AT 18.3 MC/s AND AT 100 MC/s

By C. A. SHAIN*

[*Manuscript received November 11, 1953*]

Summary

A comparison is made of the intensities of the radiation, from the general background and from the discrete sources, which have been measured at 18.3 Mc/s and at 100 Mc/s.

It is shown that the ratio of brightness temperatures at 18.3 Mc/s and 100 Mc/s is roughly constant (about 120) for different directions of observation except near the galactic equator where low values are generally observed, the ratio being a minimum near the galactic centre. Also, for the discrete sources observed at the two frequencies there is a wide scatter in the ratio of the flux densities at the two frequencies, with a possible grouping of the sources around flux density ratios of 6 and 60.

From these results the following conclusions are drawn: absorption in interstellar gas can account for the variations in the ratio of background brightness temperatures; extragalactic sources comprise a subclass of Mills's class II and may account for the polar component of the background radiation; the background radiation is not due to radiation from sources of the type so far observed.

I. INTRODUCTION

A previous paper (Shain and Higgins 1954), subsequently referred to as Paper 1, described observations of the intensity of cosmic noise at 18.3 Mc/s. During these observations a broad strip of the sky, centred on Dec. -32° , was scanned with an aerial having an overall beam width to half-power of 17° . A number of discrete sources standing out from a comparatively high intensity background was observed. A very similar but more extensive survey was made at 100 Mc/s by Bolton and Westfold (1950); with their equipment the discrete sources made no appreciable contribution to the total received power. A survey of the sky at 101 Mc/s, using the interferometer technique, to detect discrete sources has been described by Mills (1952).

It was remarked in Paper 1 that the results of the 18.3 Mc/s observations are generally similar to the results of the observations at 100 Mc/s. Nevertheless there are important differences, and the object of the present paper is to examine the similarities and differences in an attempt to throw some further light on the origin of the radiation. Comparisons of the intensities of cosmic noise as measured at different frequencies have been carried out by a number of authors, particularly Piddington (1951) and Brown and Hazard (1953). Each of these authors considered all the radiation due to discrete sources and interstellar gas, although their conclusions as to the distribution and properties of the constituents of the model galaxies differed considerably. However, all stressed

* Division of Radiophysics, C.S.I.R.O., University Grounds, Sydney.

the importance of low frequency observations and the present comparison is desirable since Paper 1 added considerably to the observational data at 18.3 Mc/s. Section II considers the general background radiation and Section III the discrete sources, while in Section IV the connexion between the discrete sources and the general background is discussed.

II. THE BACKGROUND RADIATION

In presenting the results of their 100 Mc/s survey, Bolton and Westfold (1950) gave a map of the equivalent aerial temperatures observed as their aerial scanned the sky, and they then attempted to allow for the smoothing effect of the aerial beam to obtain a closer approximation to the actual brightness distribution. In Paper 1 a map is given of the equivalent aerial temperatures observed at 18.3 Mc/s; no correction for aerial smoothing was attempted.

Since the main lobe of the aerial used for the observations of Paper 1 had nearly the same shape and size as that used by Bolton and Westfold, the smoothing of the actual brightness distribution by the aerial beams should be nearly the same in the two cases. A comparison of the *observed* distributions should then give results similar to those which would be obtained if the *actual* brightness distributions were compared, although fine details in the comparison would be obscured. However, the details of the side lobes were different for the two sets of observations and it was therefore necessary to correct the observed distributions, at both frequencies, for the different smoothing effects of side lobes.

The correction was carried out by the methods described by Bolton and Westfold, and the results, contours of equal observed brightness temperature corrected for the effects of side lobes, are given in Figure 1.*

(a) *Ratio of Observed Brightness Temperatures at 18.3 Mc/s and at 100 Mc/s*

As expected, the form of the contours in Figure 1 is generally similar at the two frequencies, apart from some irregularities on the 18.3 Mc/s contours which can be attributed to the discrete sources. However, the ratio of maximum to minimum temperature at 18.3 Mc/s is only 5.2, whereas the corresponding ratio at 100 Mc/s is 6.7. The difference, 25 per cent. of the mean, is well outside the experimental uncertainty, which in each case is not more than about ± 2 per cent. for relative intensities.

The ratio of the observed brightness temperature at 18.3 Mc/s, T_{18} , to the observed brightness temperature at 100 Mc/s, T_{100} , has been calculated over the region of the sky covered in Figure 1, and contours of equal ratio are plotted in Figure 2. Examination of Figure 2 shows that, again with some irregularities in the contours caused by the presence of discrete sources, the ratio T_{18}/T_{100} is approximately constant (within about 15 per cent.) over about 85 per cent. of the observed strip of the sky, but this ratio is generally low near the galactic

* In the original observations, different systems of galactic coordinates were used (see Paper I and also Bolton and Westfold (1951)). The coordinates of Figure 1 are based on the galactic pole at R.A. 12 hr 40 min, Dec. $+28^\circ$ (1900), and appropriate precession corrections were applied to the 100 Mc/s data.

equator from longitude 240° to at least the limit of the observed strip near longitude 350° . There is a marked dip to a ratio of 76 near the galactic centre (in the region of longitude 325° , latitude -1°). It appears that the difference in ratio of maximum to minimum brightness temperature at the two frequencies is associated with this low ratio T_{18}/T_{100} near the galactic centre.

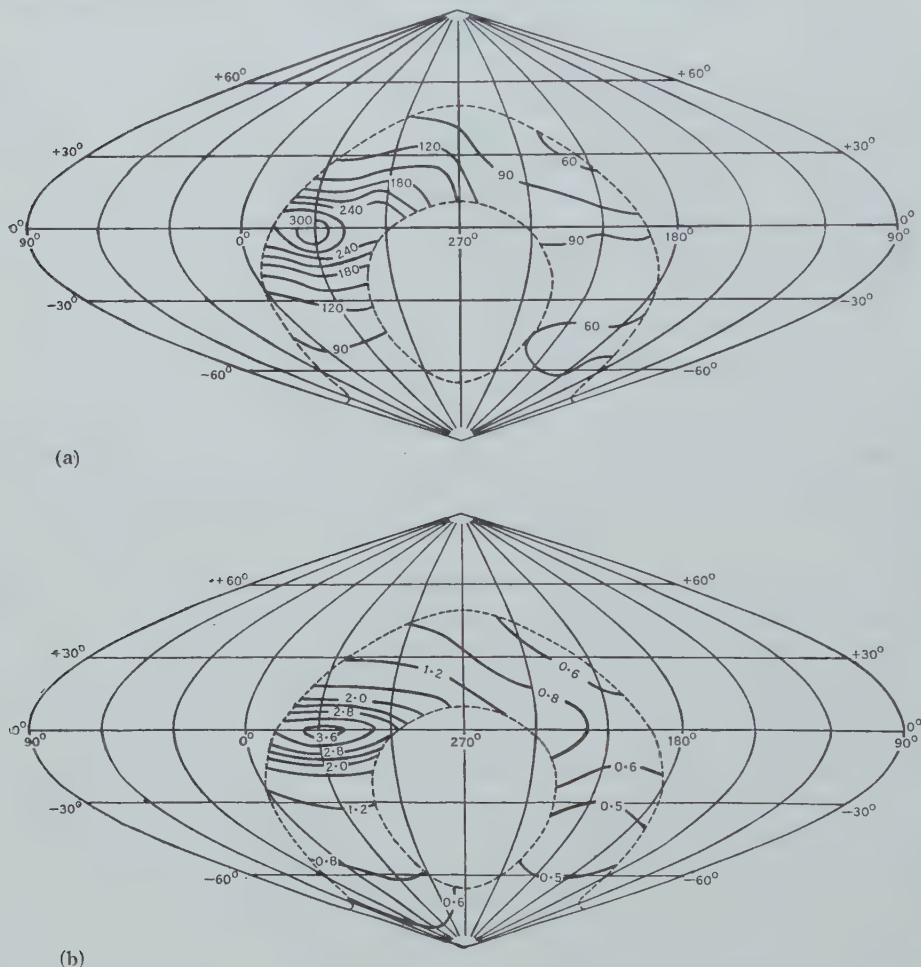


Fig. 1.—Contours of observed brightness temperatures (unit 1000 °K). (a) 18.3 Mc/s, (b) 100 Mc/s. These contours have been derived from published contours as explained in the text and in each case they correspond to the brightness temperatures which would be observed with similar idealized aerials having a single main lobe of beam width 17° .

It will be shown in Section IV that the spectra of the discrete sources differ from the spectrum of the background so that the presence of a discrete source will distort the contours in the sense of increasing the ratio T_{18}/T_{100} . To avoid complications due to this cause, two representative cross sections of Figure 2 have been drawn for restricted ranges of galactic longitude, avoiding as far as possible the known discrete sources. These sections are shown in Figure 3.

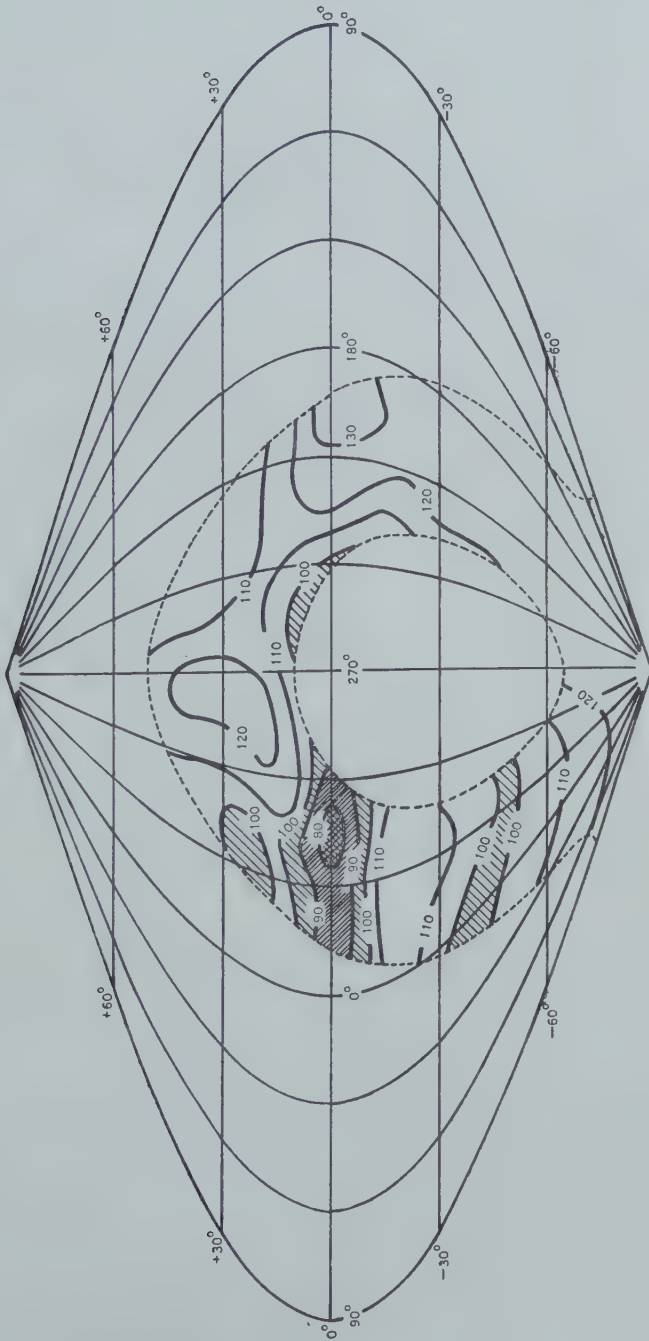


Fig. 2.—Contours of the ratio T_{18}/T_{100} derived from Figure 1.

It is seen that for longitudes between 205 and 215° the ratio is approximately constant around a mean value of about 124 , while for longitudes 305 – 330° the maximum value occurs at the pole and there is a sharp dip for latitudes less than about 20° . There are subsidiary minimum and maximum values near latitudes -50 and -25° respectively.

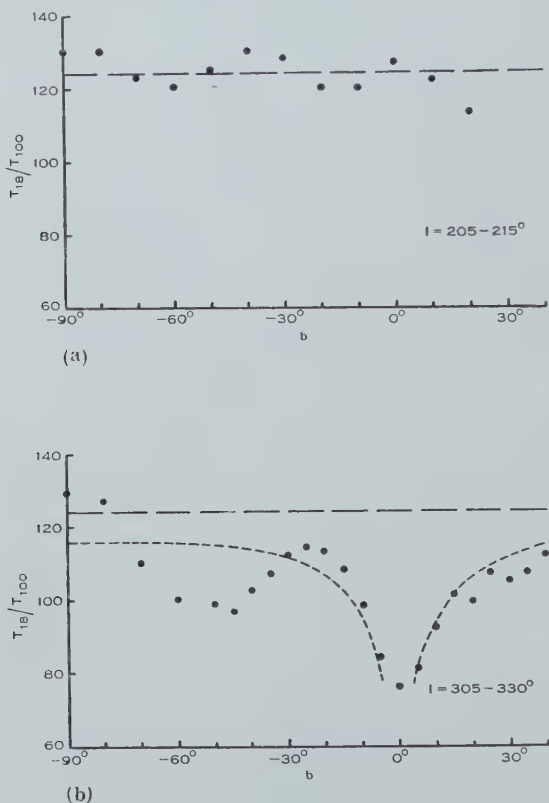


Fig. 3.—Representative sections of Figure 2, avoiding as far as possible the known discrete sources. All the points were obtained in either of two restricted ranges of galactic longitude. (a) Longitudes 205 – 215° , (b) 305 – 330° . The dashed line in (b) is a theoretical curve showing the variation to be expected due to absorption in interstellar gas.

The sharp dip in the ratio T_{18}/T_{100} near the galactic centre extends over only a fairly small range of latitudes. This suggests that it is due to some feature of the Galaxy which is confined to regions near the galactic plane. Interstellar gas is thought to extend over such a comparatively thin region near the galactic plane and, since absorption in the gas would be relatively greater at 18.3 Mc/s than at 100 Mc/s, its effect would be in the direction observed. At the frequencies concerned emission from the gas can be neglected (Westerhout and Oort 1951); therefore, as a possible explanation of the dip in T_{18}/T_{100} towards the galactic centre, the effect of interstellar absorption on the ratio T_{18}/T_{100} will now be considered.

(b) *The Effect of Interstellar Gas*

(i) *Expected Ratio of Brightness Temperatures for Two Frequencies.*—Consider the case in which interstellar gas is distributed roughly uniformly in a thin disk, symmetrical about the galactic plane. The radiation is assumed to come from some sources more widely distributed throughout the Galaxy than the gas. (The “sources” of the background radiation considered here should not be identified with the discrete sources which have been observed. It will be shown later, by consideration of the spectra of the discrete sources, that it is unlikely that these sources could account for more than a small part of the background radiation.) Let T' be the brightness temperature that would be observed, for a particular direction, in the absence of the absorbing gas. Then in directions for which the path length in the emitting region is much greater than the path length in the gas (that is, for directions more than, say, 6° from the galactic plane) it may readily be shown that the actual observed brightness temperature T is given closely by

$$T = T' e^{-\alpha h \operatorname{cosec} |b|},$$

where α is the absorption coefficient per parsec,

h is the semi-thickness of the gas disk,

b is the galactic latitude in which the observations are made.

The absorption coefficient will vary inversely as the square of the frequency f Mc/s. Hence we may put

$$\alpha = af^{-2},$$

where a will be the absorption coefficient per parsec for a frequency of 1 Mc/s. Therefore

$$T = T' \exp (-af^{-2}h \operatorname{cosec} |b|).$$

Let T_1 and T_2 be the brightness temperatures observed at frequencies f_1 and f_2 Mc/s respectively. Then

$$T_1/T_2 = (T_1'/T_2') \exp \{-ah(f_1^{-2} - f_2^{-2}) \operatorname{cosec} |b|\},$$

and, taking logarithms,

$$\ln (T_1/T_2) = \ln (T_1'/T_2') - ah(f_1^{-2} - f_2^{-2}) \operatorname{cosec} |b|.$$

To the present no restrictions have been put on the spectra of the sources of the radiation, that is, on T_1'/T_2' . We will now make the tentative assumption that all the sources have the same spectra so that the ratio T_1'/T_2' does not depend on galactic latitude. In this case a graph of $\ln (T_1/T_2)$ against $\operatorname{cosec} |b|$ should be a straight line, from which T_1'/T_2' and ah , the optical depth in the direction of the pole, may be found.

For a given direction, T_1 and T_2 are strictly the temperatures that would be observed with infinitely narrow beamed aerials whereas the observed temperatures are the averages over a fairly wide cone of angles with the axis in the direction considered. However, if only directions not too close to the galactic plane are considered, and this restriction has already been introduced, the ratio of the observed temperatures for a particular direction of the aerial maximum should be very nearly equal to T_1/T_2 for that direction.

(ii) *Comparison with Experimental Results.*—The values of the ratio T_{18}/T_{100} for galactic latitudes between 6° and 40° were taken from Figure 3 (b) and for each latitude the values of the ratio for the northern and southern hemispheres were averaged. Figure 4 shows $\ln(T_{18}/T_{100})$ plotted against $\operatorname{cosec} |b|$ and it is seen that the points lie reasonably close to a straight line. This result supports the assumptions that there is an absorbing layer in the form of a thin disk about the galactic plane and that the spectrum of the sources of the radiation is independent of galactic latitude. The dotted curve in Figure 3 (b) shows the variation of T_{18}/T_{100} with latitude corresponding to the line in Figure 4.

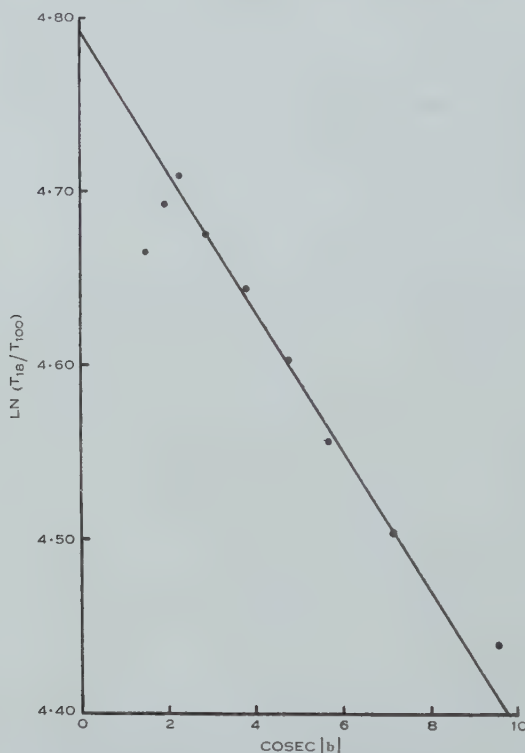


Fig. 4.— $\ln(T_{18}/T_{100})$ as a function of $\operatorname{cosec} |b|$ for the points of Figure 3 (b).

The intercept on the $\ln(T_{18}/T_{100})$ axis (Fig. 4) is 4.79, corresponding to a ratio T_{18}'/T_{100}' equal to 120, and the slope of the line is -0.04 , giving $ah=14$.

Westerhout and Oort (1951), considering the 100 Mc/s emission from interstellar gas, took the semi-thickness of the gas disk to be 100 parsecs. Also, taking the characteristics of the gas suggested by Stromgren (1948), they found an optical depth of 0.017 per kiloparsec at 100 Mc/s. This corresponds to a value for a of 0.170 per parsec, giving $ah=17$, in good agreement with the value found above, although there is considerable uncertainty in the value deduced from optical data.

From the value of ah derived above, and taking h equal to 100 parsecs, it is found that 100 Mc/s radiation in the galactic plane would be very little absorbed between the Sun and the centre of the Galaxy, whereas 18.3 Mc/s radiation would reach optical depth unity within 3 kiloparsecs. Any source of radiation near the galactic centre would be "visible" at 100 Mc/s but obscured at 18.3 Mc/s. It will be seen from Figure 1 that the maximum observed temperature at 100 Mc/s is close to the galactic plane, whilst the maximum at 18.3 Mc/s is several degrees to the south and at a different longitude. It is probable that there exists a region of high brightness near the galactic centre, perhaps the source "K" observed by Bolton *et al.* (1954), of which the brightest portion is obscured by gas absorption at 18.3 Mc/s so that at this frequency the position of the apparent maximum is displaced from the position of the 100 Mc/s maximum.

The above discussion applies to latitudes greater than about 6° . If the gas disk is only of the order of 100 parsecs thick, the radiation never travels more than a few kiloparsecs through the gas to the earth. Therefore no conclusions may be drawn as to whether the gas is in the form of a complete disk extending from the centre of the Galaxy to the Sun or in the form of an annulus with little gas in the central regions of the Galaxy. However, if the thin layer of gas extended uniformly round the Sun even for only a few kiloparsecs, there should be little change with longitude for a given latitude and in particular there should be the sharp dip near the equator for all longitudes. Examination of Figure 3 (*a*) shows that this is not so. It follows that there must be irregularities in the distribution of the gas with longitude.

In this connexion, the observations of Morgan, Sharpless, and Osterbrock (1952) are of interest. These workers found that $H\alpha$ emission regions in the northern Milky Way are confined to two lines, believed to represent two spiral arms of the Galaxy. The nearer arm extends from galactic longitude 40 to 190° and passes at its nearest point about 300 parsecs distance from the Sun in a direction opposite to that of the galactic centre. Similar observations for the southern Milky Way have been commenced by Bok, Bester, and Wade (1953), but the results for longitudes 180 – 250° have not yet been published. Christiansen and Hindman (1952), from their study of 1420 Mc/s hydrogen line radiation, also found evidence for the concentration of hydrogen to two spiral arms which would fit the results of Morgan, Sharpless, and Osterbrock.

The optical and 1420 Mc/s results confirm the generally held hypothesis that interstellar gas is mainly confined to the spiral arms of galaxies, and it is possible that in observing cosmic noise in longitude 210° we may be looking between two spiral arms, in which case there would be little interstellar gas for a considerable distance from the Sun. The absorption of 18.3 Mc/s radiation near the galactic plane for this longitude would then be small. It is probable that a similar situation occurs in visual astronomy. Bok (1937, p. 85) states that there is a notable absence of obscuring clouds of interstellar dust (generally associated with the gas) from longitudes 190 to 230° . It is also of interest that the low values of T_{18}/T_{100} in the region around $l=330^\circ$, $b=+20^\circ$ cover about the same region of the sky as the large, dark nebula in Ophiuchus. This obscuring cloud of dust extends over most of the region between longitudes 320 – 0° and

latitudes 0 to $+30^\circ$ (Bok 1937, p. 83, 1944, p. 91). Recently Sharpless and Osterbrock (1952) have detected in this region a large $H\alpha$ emission region, about 10° in diameter, near $l=330^\circ$, $b=+23^\circ$ surrounding ζ Ophiuchi. Similarly, although there is at present no optical evidence for heavy obscuration or $H\alpha$ emission in this high latitude region, the comparatively low values of T_{18}/T_{100} around $l=330^\circ$, $b=-50^\circ$ may possibly be due to the effect of a large gas cloud somewhat below the galactic plane.

If the high values of T_{18}/T_{100} near longitude 210° are associated with a lack of absorbing gas in this direction, in a gap between spiral arms, it follows that the actual sources of the background radiation cannot be confined to spiral arms. These sources must then be either extragalactic sources or galactic sources distributed more like Population II than Population I.

The analysis of the effects of absorption assumed that in the absence of absorption the ratio T_{18}/T_{100} would be the same in all directions. An alternative explanation of the variations in the observed values of T_{18}/T_{100} is that the actual sources of radiation in regions near the galactic centre have spectra intrinsically different from the sources of radiation in other regions of the Galaxy. However, although source spectral variations are not excluded, it appears that at present the variations of T_{18}/T_{100} can be best accounted for in terms of absorption in interstellar gas, the radio observations being consistent with optical evidence concerning the characteristics of the gas.

III. THE DISCRETE SOURCES

Of 37 discrete sources detected at 18.3 Mc/s and listed in Paper 1, 22 coincide in position,* within the limits of error (up to about 5°) with sources observed at 101 Mc/s by Mills (1952). These sources are listed in Table 1, with Mills's values for the galactic coordinates of each source, these being more accurate than the 18.3 Mc/s values. The sources are identified using Mills's nomenclature which gives the hour of the Right Ascension followed by the sign and tens of degrees of the declination. Sources listed in Paper 1 and by Mills are prefixed by S and M respectively. In the few cases where the designations do not correspond exactly (e.g. S03-4, M03-3), the source is close to one of the grid lines which define the designation.

In Table 1 the intensities of the sources are specified logarithmically by their "levels". The level, L_f , at a frequency f , of a source having a flux density at the earth $S_f \times 10^{-25} \text{ W m}^{-2} (\text{c/s})^{-1}$ is defined by $L_f = \log_{10} S_f$. The table gives the level of each source at 18.3 Mc/s, L_{18} , and at 101 Mc/s, L_{101} , together with the "level difference" $L_{18} - L_{101}$ which corresponds to the ratio of the flux densities at the two frequencies. For two of the sources (M08-4 and M13-4) which have appreciable angular size, Mills gave values of flux density for two spacings of the interferometer aerials. In Table 1 the greater values have been given. Bolton *et al.* (1954) have shown that the sources M13-4, M03-3, and

* The source called S20+x, which may probably be identified with the source M19+4, Cygnus-A, has been omitted owing to the great uncertainty in its position and intensity (see Paper 1).

M05-4, or perhaps sources associated with them, have considerable angular size. The total intensities measured by Bolton are such that the values of L_{101} for these sources should be increased by about 0.3, 0.4, and 0.8 respectively. The corresponding values of L_{101} and of $L_{18} - L_{101}$ are shown in brackets in Table 1. The uncertainties in the levels vary from source to source. They are generally less than 0.1 in L_{101} (corresponding to an uncertainty in S_{101} of ± 25 per cent.) and less than 0.2 in L_{18} (± 50 per cent. in S_{18}).

TABLE 1
SOURCES OBSERVED AT BOTH 18.3 Mc/s AND 101 Mc/s

Designation		Position (Mills)		L_{18}	L_{101}	Level Difference ($L_{18} - L_{101}$)
		l	b			
S01-2	M01-2	163	-73	3.0	0.9	2.1
S03-4	M03-3	206	-55	2.5	1.4 (1.8)	1.1 (0.7)
S04-2	M04-1	189	-36	2.8	1.2	1.6
S05-4	M05-4	219	-31	2.5	1.4 (2.2)	1.1 (0.3)
S06-1	M06-1	191	-17	3.0	0.8	2.2
S06-3	M06-3	206	-15	2.5	1.0	1.5
S06-5	M05-5	231	-29	2.9	1.2	1.7
S07-1	M07-2	200	-4	3.1	1.0	2.1
S08-4	M08-4	229	0	2.8	1.7	1.1
S09-1	M09-1A	212	+28	2.6	1.4	1.2
S09-3	M09-3	227	+15	2.8	0.9	1.9
S10-1	M10-0	217	+41	2.9	0.7	2.2
S12+1	M12+1	260	+74	3.0	2.1	0.9
S13-2	M13-2	282	+39	3.3	1.0	2.3
S13-4	M13-4	278	+18	3.7	2.4 (2.7)	1.3 (1.0)
S13-5	M13-6	276	0	(3.5)	1.9	(1.6)
S15-6	M16-6	293	-9	(3.8)	1.9	(1.9)
S19+0	M19+0	7	-2	3.0	1.5	1.5
S20-4	M19-5	316	-30	2.5	0.7	1.8
S21-2	M21-2	356	-53	2.7	0.9	1.8
S21-3	M21-3	341	-43	2.9	0.9	2.0
S21-5	M22-5	308	-51	2.4	0.7	1.7

A comparison of the intensities observed at the two frequencies is made in Figure 5 where, for each source listed in Table 1, L_{18} is plotted against L_{101} . For each of the sources M13-4, M03-3, and M05-4 two circles are shown corresponding to the pair of values of L_{101} . There is a considerable scatter of the points, but there is a general tendency for the intensities measured at the two frequencies to increase together.

Mills has proposed that the discrete sources be divided into two classes on the basis of spatial distribution. The class I sources would comprise intense sources distributed thinly throughout the Galaxy and hence generally close to the galactic plane; the class II sources, more or less randomly distributed, may be either weak galactic objects near the Sun or extragalactic sources. In case the spectra of these classes may be different, the sources will be divided into

two groups, one having latitudes less than 12° ("equatorial" group) and the other having latitudes greater than 12° ("non-equatorial" group). This is the division used by Mills in his analysis. In general the class I sources would be expected to fall in the equatorial group while most of the class II sources would be non-equatorial, although the separation would not be clear-cut.

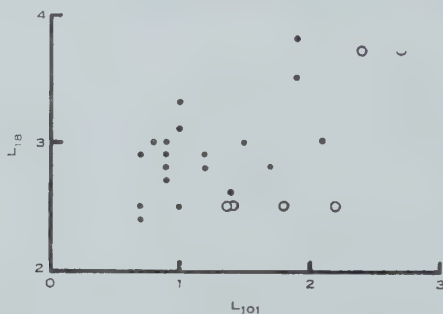


Fig. 5.— L_{18} v. L_{101} for all the sources listed in Table 1. For each of three sources two circles are given, instead of points, corresponding to values of L_{101} observed under different experimental conditions.

The values of level difference range from 0.3 to 2.3, corresponding to ratios of flux density of between 2 and 200. The variations in the spectra of the sources are shown in more detail in Figures 6 and 7. Figure 6 is a histogram of the frequency of occurrence of values of level difference. For those sources for which two values of level difference are given in Table 1, the smaller has been

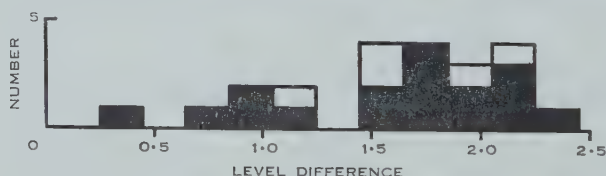


Fig. 6.—Histogram of the frequency of occurrence of level differences. The shaded portion of the histogram corresponds to the non-equatorial sources, the unshaded portion to the equatorial sources.

used. That part of the histogram which corresponds to the equatorial sources has been left unshaded, the shaded portion corresponding to the non-equatorial sources. Figure 7 shows plots of level difference against (a) L_{18} and (b) L_{101} . Circles and points correspond to the equatorial and non-equatorial sources respectively. Level difference shows no significant correlation with L_{18} , but the inverse correlation with L_{101} is statistically significant.

The histogram in Figure 6 is markedly skew. With a scatter of level differences from 0.3 to 2.3 the mode is at 1.75, the mean being at 1.6 and the standard deviation 0.5. Since the observational uncertainty in level difference would be expected to be, at most, not more than 0.3, corresponding to an

uncertainty in the ratio of flux densities of a factor of 2, it would appear that some factors other than the random errors of measuring L_{18} and L_{101} from the records must be affecting the shape of the histogram.

Two effects which must be considered as possibly affecting the shape of the histogram are angular sizes of the sources and absorption along the paths between the sources and the Earth. Besides the sources mentioned earlier, it is probable that others have angular sizes sufficiently great to give a low value of intensity with Mills's interferometer. For example, from a study of the scintillations of some sources, it was suggested in Paper 1 that the source S10-1 (M10-0) might have an angular size of between 0.5 and 1.0° . However, source size effects, which would tend to increase the values of level differences, could cause the observed skewness of the histogram only if the intensities of a moderately large

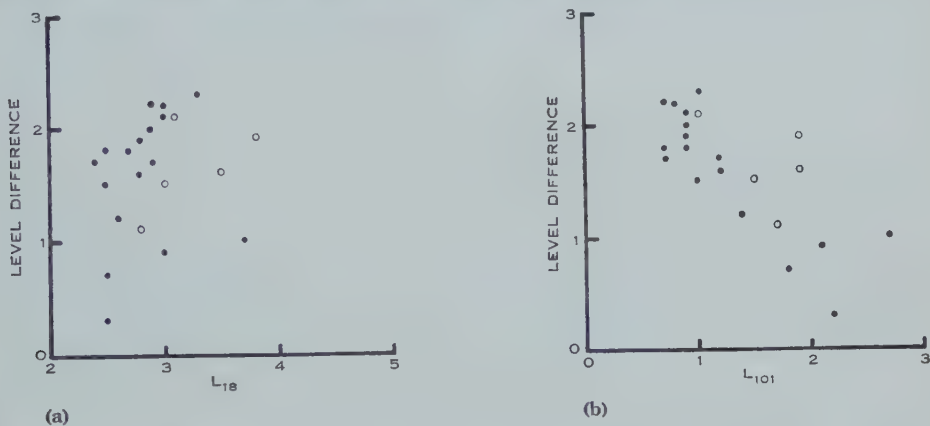


Fig. 7.—Level difference *v.* (a) L_{18} , (b) L_{101} . Points and circles relate to the non-equatorial and equatorial sources respectively.

number of sources had been affected. At present there is little evidence on which to decide whether such a possibility is reasonable, but Bolton (personal communication) considers that the percentage of sources with large angular sizes is fairly small (say 10 per cent.).

Absorption would be expected to be greatest (and hence the level differences lowest) for the sources in lowest latitudes. However, Figure 6 shows that the level differences for the sources within 12° of the galactic plane are no smaller than those for the sources outside that range of latitudes. A plot of level difference against galactic latitude (not reproduced) showed no significant correlation. Also, if absorption were markedly affecting the observed intensities of the sources, any correlation between level difference and L_{101} would be expected to be direct rather than inverse (Fig. 7 (b)). The conclusion that interstellar absorption is inappreciable for the relatively inaccurate observations of the intensities of the discrete sources is consistent with the results of the previous section.

The proposed explanation of the skewness of the histogram and of the apparent dependence of level difference on L_{101} but not on L_{18} is that these are due to differing spectral characteristics of the sources themselves. It is suggested

that the sources may be divided into two groups on the basis of their spectra, the first having level differences clustered about 0.8 and the second about 1.8, the division coming at about level difference 1.2. Five of the six sources having level differences 1.2 or less are in the non-equatorial group and this may imply a subdivision of Mills's class II into two subclasses. Two of these five sources, M12+1 and M13-4, have been identified with peculiar extragalactic nebulae, and, since none of the other sources has been so identified, the subclass having low level differences, and comprising the five strongest non-equatorial sources, may be tentatively assumed to be extragalactic objects. On this assumption it will be shown in the next section that a class of such sources could account for the general background radiation observed near the poles. The second group of non-equatorial sources have spectra not significantly different from those of the equatorial sources (level differences about 1.8). It is possible that a large group of class II sources and the class I sources are the same type of galactic objects, the more powerful emitters being concentrated towards the galactic plane.

Since the number of sources for which level differences are available is comparatively small, these suggestions must be considered as provisional until more data are available.

IV. THE DISCRETE SOURCES IN RELATION TO THE GENERAL BACKGROUND

If the general background radiation is made up of the radiation of a large number of similar discrete sources, the variation with frequency of the flux density at the Earth from an individual source should be the same as that of the brightness of the background radiation. This relation would be modified by severe interstellar absorption, but it has been shown in Section II that absorption is important only over a comparatively small area of the sky. The background brightness at a wavelength λ is proportional to $T\lambda^{-2}$ where T is the brightness temperature.

It has been shown that over most of the sky the ratio of brightness temperatures of the background at 18.3 Mc/s and at 100 Mc/s is about 120, although somewhat lower values are observed near the galactic centre. The corresponding ratio of brightness is 4. It has been shown also that the average values of the level difference for the discrete sources, taken all together, is 1.6, while if the group of sources with low level differences are omitted this would be raised to about 1.8. These level differences correspond to ratios of flux densities of 40 or 60 respectively. Although the values for some sources may be reduced as the effects of source size are overcome in further observations, it is hardly likely that these average values should be in error by a factor as great as 10; the probable errors of the averages for the sources studied correspond to a factor of 2. The variation with frequency of the source flux densities is therefore quite different from that of the background brightness and it appears that the background radiation cannot be made up of the radiation from discrete sources similar to the bulk of those so far observed.

However, the level differences for the possible extragalactic subclass of the class II sources are about 0.8, corresponding to a flux density ratio of 6. This is

close to the brightness ratio for the general background and it is possible that these sources could be responsible for at least some part of that radiation. Since the class II sources, and especially any extragalactic sources, are distributed homogeneously whereas the general background shows a marked galactic concentration, the contribution from these sources must be small. They could be responsible for the "residual" background of Bolton and Westfold (1951) and Westerhout and Oort (1951) which is the apparent reason for the high brightness temperatures observed near the galactic poles.

Mills (1952) suggested that all the class II sources might be extragalactic objects, and in some rough calculations he showed that the total radiation from all such sources would be of the right order of magnitude to account for the "residual" radiation. The restriction of Mills's calculation to the sources of the suggested subclass does not invalidate the result, since for his estimate of the density of the sources he took as the typical source M12+1, which has been included in this subclass.

V. CONCLUSION

The most important conclusion to be drawn from the comparisons of the intensities at 18.3 Mc/s and 100 Mc/s of the background and of the discrete source radiation is that the background radiation cannot be made up of the radiation from sources of the type so far observed, although a small polar component may be due to the extragalactic sources. This conclusion makes the possible identification of the sources of the background radiation even more difficult than has been thought up to the present.

It has been shown that absorption in interstellar gas must have a considerable effect on the intensity variations with direction of the background radiation observed at 18.3 Mc/s. This effect will presumably be even more pronounced in any observations made at lower frequencies. It has been also shown that the considerable scatter in the ratio of flux densities of the discrete sources at 18.3 Mc/s and at 101 Mc/s suggests a grouping of the sources in terms of their spectra, certain sources probably comprising a class of extragalactic objects. The remaining sources, including probably sources of both of Mills's classes, all have higher flux density ratios which do not differ significantly amongst themselves.

Further observations at comparatively low and high frequencies should make possible a more detailed study of the structure of the "radio Galaxy" and of the absorbing effects of interstellar gas. It is apparent that future work on the discrete sources requires the accumulation of data on flux densities and angular sizes at several frequencies for many sources so that firm statistical conclusions may be drawn concerning possible methods of classification.

VI. ACKNOWLEDGMENTS

The author is indebted to Mr. F. J. Kerr and to Mr. B. Y. Mills of the Division of Radiophysics for useful discussions during the preparation of this paper.

VII. REFERENCES

- BOK, B. J. (1937).—"The Distribution of the Stars in Space." (Univ. Chicago Press.)
- BOK, B. J. (1944).—*Pop. Astr.* **52**: 67-108.
- BOK, B. J., BESTER, M. J., and WADE, C. M. (1953).—*Sky & Telesc.* **12**: 178.
- BOLTON, J. G., and WESTFOLD, K. C. (1950).—*Aust. J. Sci. Res.* **A 3**: 19-33.
- BOLTON, J. G., and WESTFOLD, K. C. (1951).—*Aust. J. Sci. Res.* **A 4**: 476-88.
- BOLTON, J. G., WESTFOLD, K. C., STANLEY, G. J., and SLEE, O. B. (1954).—*Aust. J. Phys.* **7**: 96.
- BROWN, R. H., and HAZARD, C. (1953).—*Phil. Mag.* **44**: 939-63.
- CHRISTIANSEN, W. N., and HINDMAN, J. V. (1952).—*Aust. J. Sci. Res.* **A 5**: 437-55.
- MILLS, B. Y. (1952).—*Aust. J. Sci. Res.* **A 5**: 266-87.
- MORGAN, W. W., SHARPLESS, S., and OSTERBROCK, D. (1952).—*Sky & Telesc.* **11**: 138-9.
- PIDDINGTON, J. H. (1951).—*Mon. Not. R. Astr. Soc.* **111**: 45-63.
- SHAIN, C. A., and HIGGINS, C. S. (1954).—*Aust. J. Phys.* **7**: 130.
- SHARPLESS, S., and OSTERBROCK, D. (1952).—*Astrophys. J.* **115**: 89-93.
- STROMGREN, B. (1948).—*Astrophys. J.* **108**: 242-75.
- WESTERHOUT, G., and OORT, J. H. (1951).—*B.A.N.* **11**: 323-33.

HIGH MULTIPLE RADIO REFLECTIONS FROM THE F_2 LAYER OF THE IONOSPHERE AT BRISBANE

By K. BAIRD*

[*Manuscript received October 16, 1953*]

Summary

Continuous night-time records of multiple F_2 reflections at normal incidence have been made at a fixed frequency. The echo patterns have been classified, and qualitative explanations given in terms of humped ionization contours, extending the work of Pierce and Mimno (1940). These patterns have been studied also by a variable gain technique. It is concluded that accurate measurements of reflection coefficients cannot be made by this means. Statistical analyses of occurrences of up to the 10th multiple showed that: (a) if no account is taken of presence or absence of E_s , the frequency of occurrence increases towards dawn; (b) there is no correlation between the number of reflections observed and the virtual height of the region; (c) there is no correlation with "range duplications"; (d) inverse correlation between high multiple F reflections and presence of E_s occurs only when the lower region is blanketing; (e) there is no correlation between high multiples and the travelling disturbances described by Munro.

A study of the presence of very high multiples revealed maxima at the equinoxes.

Oblique incidence recording gave no reflections beyond the fifth multiple.

I. INTRODUCTION

Part of this investigation is an extension of the work done by Pierce and Mimno (1940) on the reception of radio waves from distant ionospheric irregularities. Their work was carried out at a frequency of 3.5 Mc/s, and they proposed a plausible mechanism by which high multiple echo patterns could be formed.

The present work also includes a statistical study of the occurrences of high multiples, and of their correlation with other ionospheric phenomena, such as sporadic E region, range duplications,† and cellular waves.

II. EQUIPMENT

The equipment operated at a frequency of 2.28 Mc/s, using a pulse transmitter of peak output about 1 kW. Its pulse repetition frequency was 50 per second, and pulse width, 70 μ sec. The receiver was a modified Loran A/APN4 type, of sensitivity requiring about 50 μ V at the receiver input for recording of an echo. Since the aerial used was a half-wave dipole, this is equivalent to a field strength of about 2 μ V/m.

* Physics Department, University of Queensland; present address: Physics and Engineering Unit, Wool Textile Research Laboratory, C.S.I.R.O., Sydney.

† A satellite echo above F_2 , giving on $p'f$ records a trace approximately duplicating F_2 , except in equivalent range (see Section X below).

The time base of the display system used for slightly more than half the investigation was adjusted to allow echoes of equivalent path up to 5600 km to be observed; for the remainder of the time, a path length up to 8000 km was recorded. Thus 16 reflections from a layer of virtual height 250 km could just be recorded in the latter case. All recording employed a camera using 35-mm film moving past the image of the 6-in. cathode-ray oscillograph screen at a rate of 5 in/hr.

As there is no F_2 reflection during the day at this frequency, the records are for night-time only, the equipment being run for up to 10 hr each night.

III. ECHO PATTERNS OBTAINED

Examples of the various echo patterns obtained using vertical propagation, and simplified sketches of them, are shown in columns *A* and *B* of Plate 1. Time and height scales on the drawings are approximate only, as these are variable.

On examination of 200 nights' recording, the following numbers of the various patterns were observed:

(a) or (b) 155; (c) 2; (d) 1; (e) 2; (f) 5; (g) 3; (h) 2.

On a few occasions, somewhat similar patterns of multiple E_s reflections have been noticed, but the records have not been fully examined for such occurrences.

IV. DISCUSSION OF ECHO PATTERNS

The assumption of one simple ionospheric irregularity, with minor variations, is sufficient to give a qualitative explanation of all the echo patterns of Plate 1 (a)–(f).

Following Pierce and Mimno (1940) suppose that a hump, such as shown in Figure 1 (a), moves over the transmitter-receiver site, where the angle α is about 1° , and the overall extent may be 200 km. This is an over-simplification, but at such small curvatures, straight-line sections are a reasonable approximation. Assuming geometrical optics to be applicable, a ray can retrace its path and be returned to the receiver only if it somewhere strikes a surface normally. Therefore only those rays leaving the transmitter at an angle $n\alpha$ (where n is an integer) to the normal can be returned. For in *AB* each reflection at the ionosphere or ground increases the angle between ray and normal by α , and in *BC* a corresponding decrease of angle α occurs at each reflection.

The limited extent of *AB* and *BC* imposes further conditions, so that for any particular starting point, reflection occurs only for certain values of n . The echo pattern shown in Figure 1 (b) was derived by calculating all the paths of the rays for a series of positions of the irregularity, displaced horizontally with respect to each other.

If we regard the two sections *AB* and *BC* of Figure 1 (a) as separate entities, and combine them in the manner indicated in column *C* of Plate 1, we obtain by similar means exactly the patterns shown in column *B*. Now, changes in the irregularity while passing overhead, asymmetry in the irregularity, changes in angle, height, extent, or speed of the movement will all cause modifications to the patterns considered; all the echo patterns shown in column *A* of Plate 1 (a)–(f) can thus be accounted for.

It was noted that on many nights, especially in midsummer, an echo pattern of type (a) of Plate 1 occurred just before dawn. (This occurrence causes a maximum at this time in the histograms of diurnal variation (Fig. 2).) The pattern obtained did not quite agree with those described by Pierce and Mimno, but was more often of the type here shown. This is exactly the type of formation which would be expected from a region of large extent having about 1° inclination to the horizontal, and moving towards the recording site. This is precisely the state of affairs at this time of day ; shortly afterwards, the F_2 reflection disappears altogether.

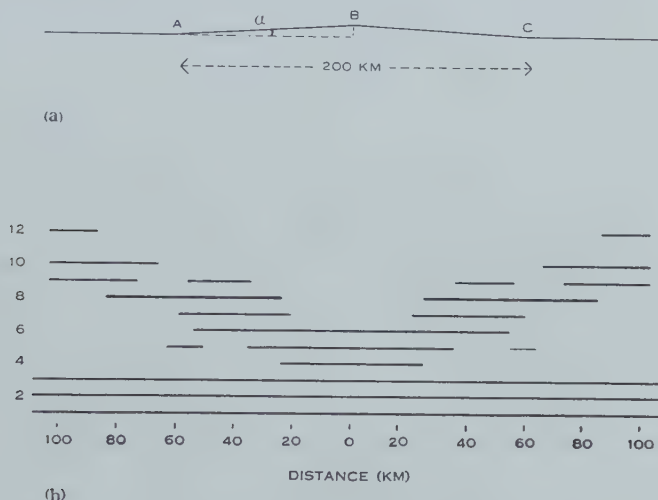


Fig. 1.—Echo pattern formed by ionospheric hump.

(a) Assumed irregularity in the F_2 region.

(b) Echo pattern calculated from (a).

It is worth noting that the very simple irregularity assumed in Figure 1 gives rise to a somewhat complex echo pattern—there are some orders missing, some are intermittent, while others appear very strongly. Thus it is to be expected that recorded echo patterns would often be very complex, considering the complicated horizontal distribution of electron density which certainly occurs in the F_2 region.

The pattern type (g) will be dealt with in Section IX.

The pattern (h) is quite different in appearance from (a)–(f). It can be explained along the following lines.

If the angle α (Fig. 1) is increased to about 3° , and the overall extent of the hump decreased to about 100 km, the geometry changes, and patterns such as those of Plate 1 (a)–(f) cannot be formed. However, with a configuration differing from this only by having smoothed corners, the number of multiples increases markedly as the centre of the irregularity approaches a position vertically above the transmitter-receiver site. An echo pattern of the type (h) can thus be produced. This hypothesis finds confirmation in the slight height rise shown in the records towards the centre of this pattern ; other high multiple

patterns are not accompanied by observable height rises, probably because these are below the resolution of the equipment.

Type (*k*) occurs for some 2 or 3 hr at a time and, on a night when such a reflection is seen, the other reflection patterns mentioned above are absent. The simplest explanation of such an occurrence is in terms of a very "quiet" F_2 region, which remains parallel to the Earth's surface and highly reflecting for long periods.

Thus it appears that simple kinks in the F_2 region, of various angles and widths, are adequate for satisfactory explanations of all observed high multiple patterns.

V. MEASUREMENT OF ABSORPTION

Some observations were made using a variable gain technique,* in which the receiver gain was varied from a maximum to a minimum once every minute. The total attenuation from one limit to the other was about 100 db. As before, the records are on 35-mm film which moved continuously past the image of the cathode-ray oscillograph screen at about 5 in/hr. The gaps between short sections of the trace are caused by the receiver running at minimum gain for a few seconds before returning to the maximum value. An example of the records is shown in Plate 2, Figure 1.

The lengths of the traces in any small segment indicate directly the relative intensity of an echo in decibels above the limit of recording. Owing to fading and the limitations of the recording technique, an accuracy in determination of the attenuation per reflection better than ± 5 per cent. for small attenuation, and ± 10 per cent. for large attenuation, cannot be achieved.

Table 1 (*a*) shows some readings obtained by this method, at random times during "steady" conditions when none of the echo patterns (*a*)–(*g*) of Plate 1 was occurring. This set of results illustrates the errors incurred in measuring absorption losses by taking the relative strengths of consecutive echoes without careful selection of data; for in several cases an echo of high order is stronger than those immediately below it. This would lead to a reflection coefficient greater than unity.

However, this difficulty seldom occurs in the lower orders, and a good indication of the reflection losses can be obtained by taking account of the relative strengths of the first three or four echoes. Allowing for the attenuation suffered because of increase in path length, it is found that, when high multiples are occurring, the loss per reflection is about 1–2 db.

Table 1 (*b*) shows a continuous set of results during an echo pattern of type (*a*) of Plate 1, readings being taken at 3-min intervals, as shown. The variations in intensity are fully consistent with the explanation outlined in Section IV.

* Suggested by Webster, and developed by McNicol, for another purpose.

VI. DIURNAL VARIATION

The diurnal variation of high multiple occurrence has been studied in two ways.

(a) Each night was divided into 10-min periods. When a multiple above the eighth (as an arbitrary standard) appeared, a single occurrence was recorded. For multiples between fourth and eighth, only a half was noted. Figure 2 shows the resulting histogram.

TABLE I
VALUES OF ATTENUATION OF RECEIVED ECHOES, MEASURED
FROM HIGH MULTIPLE RECORDS OF VARIABLE GAIN TYPE

(a) *Values at random times*

Multiple No.	Attenuation (db)						
11	—	—	—	—	—	—	—
10	—	—	56	41	63	71	63
9	—	—	50	50	68	61	63
8	—	70	42	57	70	59	51
7	—	88	52	69	72	63	51
6	69	76	48	66	70	73	51
5	67	70	44	48	63	71	49
4	59	52	42	50	59	60	44
3	49	61	38	41	53	54	41
2	44	55	30	41	53	44	34
1	34	36	33	41	31	35	39

(b) *Consecutive values during a period of 54 min*

Multiple No.	Attenuation (db)																		
16	60	50	40	60	60	70	70	70	70	75	—	—	—	—	—	—	—	—	—
15	70	50	50	50	60	60	60	60	60	70	70	—	—	—	—	—	—	—	—
14	—	50	50	50	60	50	60	60	60	60	60	—	—	—	—	—	—	—	—
13	—	—	50	60	50	50	50	40	40	50	50	70	—	—	—	—	—	—	—
12	—	—	60	60	60	40	40	40	40	30	40	60	—	—	—	—	—	—	—
11	—	—	—	75	70	40	40	35	30	25	40	50	75	75	—	—	—	—	—
10	—	—	—	—	75	75	60	40	30	30	40	40	70	70	75	75	—	—	—
9	—	—	—	—	—	—	70	60	30	25	40	40	70	70	70	75	—	—	—
8	—	—	75	75	75	75	70	60	40	40	40	40	70	60	50	60	70	70	75
7	—	—	70	75	75	75	60	50	40	40	40	40	60	30	40	50	60	60	70
6	70	60	60	60	60	60	60	40	40	30	30	30	40	40	40	40	40	50	50
5	50	40	50	50	50	50	50	35	40	30	30	30	30	30	30	40	30	40	40
4	35	40	40	40	40	40	40	25	35	25	25	30	40	25	25	30	15	30	30
3	35	25	25	30	30	35	25	25	25	15	15	15	25	25	15	25	25	25	25
2	25	20	15	25	25	25	15	25	15	10	10	10	15	15	15	15	10	15	15
1	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10
Time	0507	0510	0513	0516	0519	0522	0525	0528	0531	0534	0537	0540	0543	0546	0549	0552	0555	0558	0601

There is a tendency for the number of occurrences to be fewer earlier in the night. This may be due to more frequent blanketing by E_s between 1900 and 2400 hr, often preventing any F_2 reflection at all.

(b) For each hour of the night, the highest multiple showing on the records* was noted. The numbers were averaged for each hour of the night over a period of 30 days. The results are given in Table 2; they show no consistent trend.

VII. OCCURRENCE OF VERY HIGH MULTIPLES FROM $p'f$ RECORDS

Very high multiple reflections occasionally produce records on the third time base sweep of the standard variable frequency records† made at 10 min. intervals. The equivalent range is therefore somewhere over 6000 km. The order of reflection is about 20, depending on the virtual height of the region.

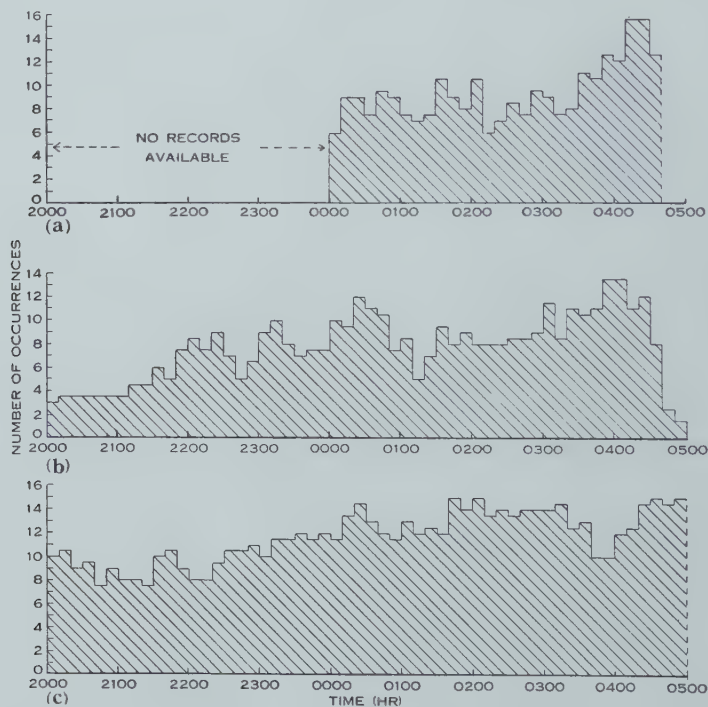


Fig. 2.—Diurnal variation of high multiple occurrences: one for multiples over the eighth, one-half for multiples between fourth and eighth.

(a) For 30 days; October 1951.

(b) For 60 days; November, December 1951 and January 1952.

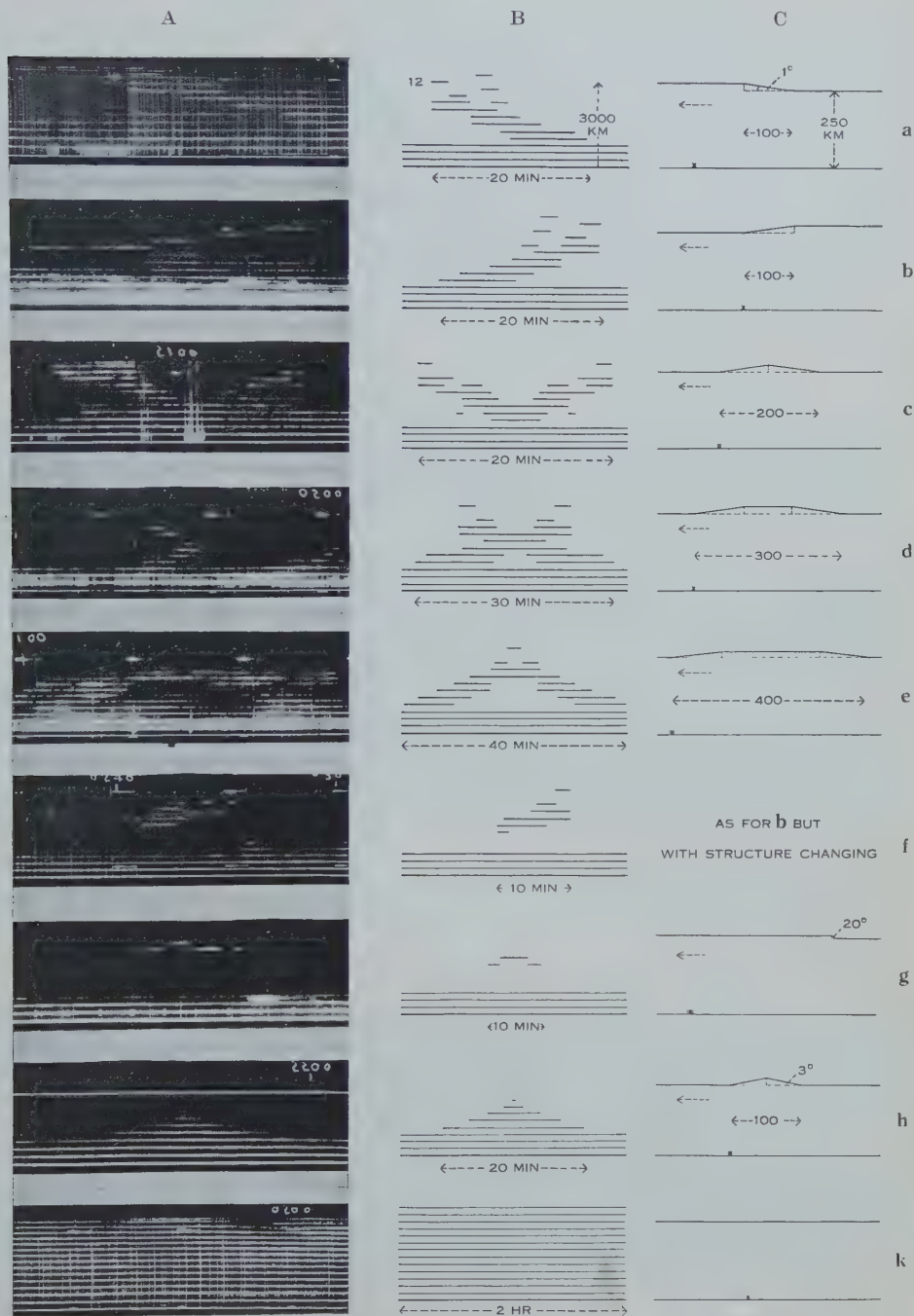
(c) For 38 days; March 1952 and some of April 1952.

The actual order for a particular record was calculated in two ways:

(a) by comparing the tangents of the angles which the first reflection trace and the n th reflection trace make with the horizontal;

* Since the display system eliminated echoes above about the 15th, these numbers were in some instances under-estimates.

† For a description of the recorder see Higgs (1945).

REFLECTIONS FROM THE F₂ LAYER OF THE IONOSPHERE

Echo patterns and their interpretations. *A*, Actual records; *B*, calculated patterns; *C*, corresponding ionospheric irregularities assumed for calculation of *B*.

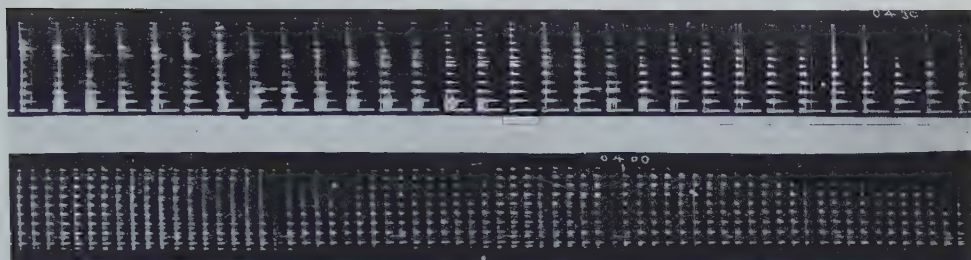
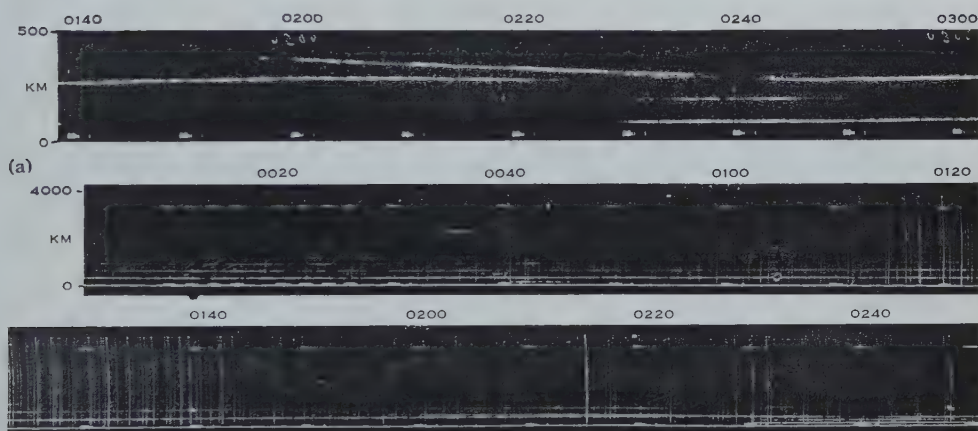
REFLECTIONS FROM THE F_2 LAYER OF THE IONOSPHERE

Fig. 1.—Records obtained with variable gain technique



(b)

Fig. 2.—The fixed frequency records during a range duplication. (a) 0–500 km film, (b) 0–4000 km high multiple film. (The periodic marks near the bottom of film (a) and near the top of film (b) occur at 10-min intervals, and are caused by the standard variable-frequency recorder.)



Fig. 3.—High multiple and 0–500 km records during E_s movement. (As in Figure 2, the p/f marks occur at 10-min intervals.)

(b) by dividing (6000 + the height of the n th reflection as shown on the record) by the height of the first reflection at that frequency.

These two methods should agree.

Records were examined for every night during 1951 and 1952, this period containing some 40,000 individual records. Extra-high multiples showed on

TABLE 2
THE MAXIMUM NUMBER OF MULTIPLES OCCURRING IN EACH HOUR OF THE NIGHT, AVERAGED FOR 30 DAYS AT A TIME

Time Period	1900	2000	2100	2200	2300	0000	0100	0200	0300	0400	0500
October 1951 ..	—	—	—	—	—	7	7.5	7.5	6.5	7.5	
January 1952	5	4.5	5	6	6.5	6	6	5.5	6.5	6.5	
March 1952 ..	9	8	9	9	9	10	10	10	9	9	

only 35 occasions. (For comparison, it should be pointed out that high multiples of approximately the 10th order were recorded, on an average, several times each night. This agrees with the frequency of occurrence of such multiples on the fixed frequency records.) The number of occurrences is plotted as a function of

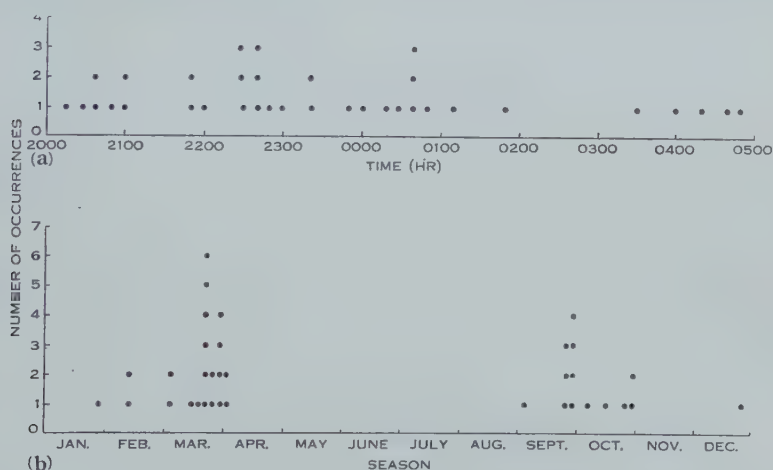


Fig. 3.—Distribution of very high multiple occurrences for 2 years, 1951 and 1952.

(a) Diurnal, (b) seasonal.

local time in Figure 3 (a). Considering the small number involved, it can be said only that extra-high multiples are more likely to occur before 0100 hr than between then and sunrise.

Their seasonal distribution, irrespective of the hour of occurrence, is shown in Figure 3 (b). The small number again precludes any definite conclusion, but the grouping suggests marked maxima at the equinoxes.

VIII. CORRELATION BETWEEN VIRTUAL HEIGHT OF THE LAYER AND NUMBER OF MULTIPLES

The results of seven nights' observations were examined in two ways :

- (a) plotting the layer height and multiple number individually as functions of time throughout the night, and comparing the graphs ;
- (b) plotting the number of multiples as a function of layer height, irrespective of time.

Both methods indicated that the number of multiple reflections is quite independent of layer height.

IX. CORRELATION BETWEEN RANGE DUPLICATIONS AND HIGH MULTIPLES

The term "range duplication" has been introduced by McNicol (personal communication 1952) to describe the phenomenon in which the F_2 echo is accompanied by a satellite,* this satellite persisting and changing in range in the same way as the main F_2 echo, as the transmitter frequency is changed. The critical frequency of the satellite trace on the $p'f$ record is normally (but not always) the same as for the principal trace. Echoes such as $F+E_s$ and $2F-E_s$ are, of course, excluded.

On fixed frequency records, the trace of lower range is always the continuous one ; the higher range trace may persist for periods up to 2 hr. Three general types of satellite are recognized :

(a) "Outgoing." The satellite starts at the same virtual range as the principal echo and gradually increases in separation. See Figure 4 (a).

(b) "Incoming." The satellite starts above the main trace and decreases in virtual range until it joins (or nearly joins) the principal trace, as in Figure 4 (b).

(c) "Bilateral." The satellite exhibits both (b) and (a) characteristics. See Figure 4 (c).

In connexion with the extensive work being carried out at Brisbane on this phenomenon, the records for 22 definite range duplications were investigated in two ways :

(a) the maximum number of high multiple reflections in a 10-min interval was plotted as a function of the time for 2 hr before and after the bifurcation point of the range duplication ;

(b) the hourly average numbers of multiples for 2 hr before and after the bifurcation were tabulated.

Neither method revealed any consistent feature. It therefore appears that the ionospheric configuration causing range duplications does not normally give rise to high multiple reflections, probably because range duplications are produced by an irregularity having a small extent (a few kilometres only) and a large inclination to the horizontal, while the patterns previously discussed require smaller inclinations.

* See also Gipps, Gipps, and Venton (1948).

Calculations by the methods outlined in Section IV indicate, however, that such a small kink in the F_2 region, while it would not lead to any high multiples when close to the recording site, would, as it approaches from a large distance, pass through one short section in which it could return an isolated high multiple echo, as pictured in Plate 1 (*g*).

A search for such isolated high multiples at appropriate times before and after the bifurcation of a range duplication was successful in one instance. This is pictured in Plate 2, Figure 2. In this case the eighth multiple appears



Fig. 4.—Idealized appearance of range duplications on fixed frequency records.

(a) "Outgoing", (b) "incoming", (c) "bilateral".

for a brief period about 2 hr prior to the junction of the satellite trace and the principal trace. In this instance, all the facts can be fitted by a kink having an inclination of 20° to the horizontal, and an extent of 3–4 km in the direction of travel. For such a kink, the vertical extent would be at most 1.5 km so that the height change would not be detectable in the records. Calculation of the

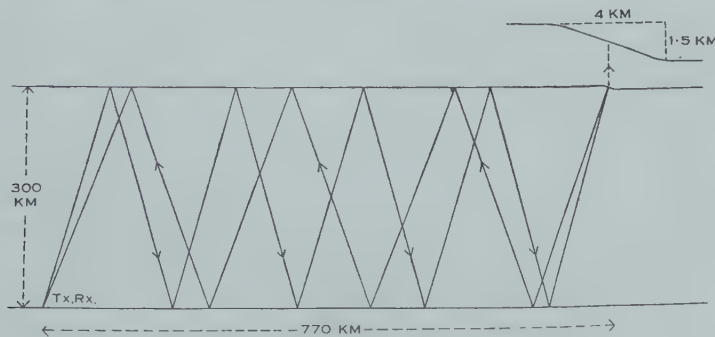


Fig. 5.—Mechanism of formation of a high multiple from a kink producing a range duplication.

velocity from the range duplication permits the placing of the kink 2 hr before the junction at 770 km horizontal displacement. At this distance (and with an angle of 20°) the eighth multiple reflection should appear, as in Figure 5.

On this view, the failure of an isolated multiple to appear near other range duplications can be ascribed to the limited stability of the ionospheric configuration. However, it is surprising that records of type (*g*) appear so rarely.

X. HIGH MULTIPLE F REFLECTION AND SPORADIC E

The occurrence of high multiple reflections is affected by the presence of E_s only when it is blanketing the upper layer, either totally or partially. In this case the strength of all F reflections is reduced by attenuation in the lower layer.

At other times, reflections from the E_s can appear in or disappear from the records without causing any noticeable change in the high multiples. An example of this is given in Plate 2, Figure 3, together with the 0–500 km record showing details of the E_s movement.

XI. CORRELATION BETWEEN HIGH MULTIPLES AND TRAVELLING DISTURBANCES

Munro (1950) has described a type of disturbance in $p'f$ records which Martyn (1950) has explained in terms of cellular waves in the ionosphere. Each occurrence of a disturbance of this type is noted by the observer scaling the Brisbane $p'f$ records.

The criterion taken for such a disturbance is the appearance of an upward kink in the height of the trace, appearing first at the higher frequency end. This then moves towards lower frequencies, causing a rise in minimum height. Also the o and x rays are separated to a much lower frequency than in the normal records preceding and following. The magnitude of the height rise is 50–60 km, and the duration, about 30 min.

Seventeen occurred during the currency of high multiple records. As for range duplications, plots were made of numbers of multiples for 2 hr before and after the passage of the disturbance, and the hourly averages for 2 hr before and after the disturbances were recorded. These show no consistent feature, indicating that the high multiples and Munro disturbances do not arise simultaneously from a given irregularity.

XII. OBLIQUE INCIDENCE

For a short period, simultaneous records were made at Brisbane of echoes arising from two transmitters, one located at the same site, and the other at Toowoomba, 100 km away. In the latter case the calculated angle of incidence at the F_2 layer, for a single hop, was about 10° .

In 25 nights' useful recording at oblique incidence, the highest multiple recorded was the fifth, which persisted for a few minutes on three occasions. The fourth was observed on 10 nights, the third and second much more frequently. During this period the normal incidence recording frequently showed multiples above the 10th reflection. This absence of oblique incidence high multiple reflections agrees with the statement made by Pierce and Mimno (1940).

There is a general correspondence between the normal and oblique records, however, the latter showing its higher multiple reflection at times when the former exhibited highest multiples and strongest echoes.

XIII. ACKNOWLEDGMENTS

The author is deeply indebted to Professor H. C. Webster, who suggested the project and maintained a very active interest in it; and to Mr. R. W. E. McNicol, without whose guidance and many helpful ideas the work could not have been accomplished.

This investigation forms part of the programme of the Radio Research Board of C.S.I.R.O. and is published by permission of the Board.

XIV. REFERENCES

- GIPPS, G. DE V., GIPPS, D. I., and VENTON, H. R. (1948).—*J. Coun. Sci. Industr. Res. Aust.* **21**: 215.
HIGGS, A. J. (1945).—Coun. Sci. Industr. Res. Aust. Div. Radiophysics Rep. PD252.
MARTYN, D. F. (1950).—*Proc. Roy. Soc. A* **201**: 216.
MUNRO, G. H. (1950).—*Proc. Roy. Soc. A* **202**: 208.
PIERCE, J. A., and MIMNO, H. R. (1940).—*Phys. Rev.* **57**: 95.

CONVECTION FROM A LARGE HORIZONTAL SURFACE

By C. H. B. PRIESTLEY*

[Manuscript received October 29, 1953]

Summary

A theory is put forward for convection from a large horizontal heated surface in a semi-infinite medium, by buoyant elements which are subject to continuous mixing with the environment by turbulence on a smaller scale. It is assumed, with support from similarity arguments, that the (potential) temperature θ at sufficient height z above the surface obeys the form $-(g/\theta)(\partial\theta/\partial z) = Cz^{-\delta}$, where δ and C are positive constants. It is then shown: (i) that δ must in practice be close to $4/3$ and equal to it under steady conditions, except in layers where radiational heating is large, where δ will be smaller; (ii) that the rate of heat loss varies as $C^{3/2}$; and (iii) that the r.m.s. temperature fluctuations are proportional to $Cz^{-1/3}$. Experimental results from the surface layers of atmosphere support these predictions quite well.

The principal results are first suggested for free convection by dimensional and similarity arguments. They receive independent confirmation from the mechanistic theory, which extends into conditions when forced convection is present but not dominant. The theory also provides information about the multiplying constants in the above relationships, though it does not so far lead to a prediction of their exact values. The multiplying constants depend, *inter alia*, on the mass ratio between the ascending and descending air, and this remains constant through the layer of constant heat flux. The behaviour of the ascending elements and that of the descending air are shown to be quite differently governed.

I. INTRODUCTION

The problem of free convection is commonly approached on empirical lines working on a foundation of dimensional analysis. Accounts of the work appear in textbooks by Fishenden and Saunders (1932), Bosworth (1952), and others, and in the meteorological literature by Sutton (1953). The formulation is in terms of three dimensionless bulk parameters, the Nusselt, Grashof, and Prandtl numbers, and formulae for heat transfer by free convection are in general suitably expressed as a functional relation between these three numbers. Experiments on convection from heated wires and pipes and vertical planes have led to the common result that the heat loss is proportional to the $5/4$ power of the temperature difference between the heated body and the surrounding medium, the index changing to $4/3$ when the motion becomes sufficiently turbulent. Other experiments (Chandra 1938; de Graaf and van der Held 1953) relate to convection between parallel planes. In both these types of study, the formation of large buoyant elements is inhibited by the scale and geometry of the experiment.

* Section of Meteorological Physics, C.S.I.R.O., Melbourne.

Special considerations must apply to convection from large (theoretically infinite) horizontal surfaces, in a medium of large vertical extent, which is the case to be considered here. The temperature difference $\Delta\theta$ between body and surroundings is no longer clearly defined since it will continue to increase, though at a continuously decreasing rate, as the height z above the surface increases. Secondly, the heated body ceases to have an identifiable linear dimension d , which is required for the specification of both the Nusselt and Grashof numbers. With bodies of finite size this length d enters largely as a factor determining the size and structure of the significant heat-carrying elements. For an unlimited heated body and medium, the size of these elements becomes a function of z and the problem becomes one-dimensional, the statistical properties of the flow and temperature fields being dependent on z but not on any horizontal coordinate. It is evident that in some respects z must replace d as one of the basic independent variables of the problem and that the Grashof number, which contains both d and $\Delta\theta$ as constituents, must lose its character as a uniquely defined bulk parameter. It is possible to employ a formally similar number with z replacing d and $\Delta\theta$ recognized as being dependent on z , but a number so defined can only be of local significance; similar considerations apply to the Nusselt number and with these changes of character it seems desirable to approach the problem afresh.

When the heated surface is large enough, it may be supposed that the main buoyant elements at sufficient height are so large that ranges of smaller-scale motions are interposed between them and the molecular scale; the mixing between these elements and their environment will itself be a turbulent type of process. This has been recognized in an earlier treatment by G. I. Taylor, described by Sutton (1953). Very close to the ground there is no scope for the full development of the sub-elemental turbulence, so that the assumption above can only be valid above a certain layer of thickness Δ ; consideration will be restricted to $z > \Delta$. In conformity, $\Delta\theta$ must relate to the temperature difference between the level under consideration and that at $z = \Delta$ rather than at the surface itself.

The problem will be subjected to dimensional analysis in Section II in order to indicate the form of relations to be expected. Following the type of postulate commonly made in modern theories of turbulence, it will here be assumed that the mixing of the elements and the consequences of that mixing are not directly dependent on the molecular constants (thermal conductivity and viscosity) of the medium in which the convection takes place. In the remaining sections a more mechanistic theory will be formulated and checked against experimental data. The concern here is pre-eminently with the meteorological problem of convection from the heated ground or water surface but the treatment should permit of considerably wider application.

II. DIMENSIONAL ANALYSIS

Given the geometry of a problem in free convection, it is known (e.g. Bosworth, loc. cit.) that six independent quantities are required to specify the physical condition, and the task is to express any dependent quantity, notably

the rate of heat loss, in terms of these six. From purely dimensional arguments the solution is not determinate but admits of two degrees of freedom.

In the present problem, the assumption concerning the scale of convective elements effectively eliminates two of the independent variables, and the form of the solution becomes determinate on dimensional grounds alone. It is taken that the heat flux F through the level z will in general depend on z and on the (potential) temperature difference $\Delta\theta$ between the levels z and Δ , on the specific heat per unit volume, c , and on gravity, which exerts its influence through the effect of thermal expansion and so for a perfect gas appears as g/T , where T is the absolute temperature. Putting then

$$F \propto z^\alpha (\Delta\theta)^\beta \left(\frac{g}{T}\right)^\gamma c^\delta,$$

and equating dimensions the required relation is derived as

$$F \propto c \left(\frac{gz}{T}\right)^{1/2} (\Delta\theta)^{3/2} \dots\dots\dots (1)^*$$

This law contrasts with the experimental $(\Delta\theta)^{5/4}$ and (turbulent) $(\Delta\theta)^{4/3}$ laws for small heated bodies.

It has been implied that the assignment of a value $\Delta\theta$ at a single specified height z is sufficient to specify the temperature state; were this not so, the result (1) should contain an additional factor

$$f\left(\frac{z}{\Delta\theta} \frac{\partial\theta}{\partial z}, z \frac{\partial^2\theta}{\partial z^2} \left/ \frac{\partial\theta}{\partial z}, \dots \right.\right),$$

where f is an arbitrary function of the non-dimensional characteristics of the temperature profile which appear within the bracket. Now it has been shown more generally by Batchelor (1953) that the condition for similarity of two flow patterns of the type considered here is that their Richardson numbers Ri shall be equal. The dimensionless profile characteristics are therefore functions solely of Ri . In the special case of free convection $Ri = -\infty$, the characteristics must be constants and the supposition that the profile is specified by a single value $\Delta\theta$ at a stated height z is confirmed. As a corollary, since, for example, $z(\partial^2\theta/\partial z^2)/(\partial\theta/\partial z)$ is constant with height, the temperature profile will take the form

$$\frac{\partial\theta}{\partial z} \propto z^{-\delta},$$

where δ is a constant, a form which has been anticipated in the literature (Sutton 1953).

* Using the Grashof and Nusselt numbers in the modified sense discussed above, (1) may be written

$$(Nu)^2 (Gr)^{-1} (Pr)^{-2} = \text{const.}$$

This represents the only relation between these three numbers which involves neither of the molecular coefficients.

Returning to equation (1), in the special case where the heat flux is constant with height and provided that, as in the meteorological problem, the absolute temperature varies only within narrow limits, it follows that

$$\Delta\theta \propto z^{-1/3},$$

or

$$\frac{\partial\theta}{\partial z} \propto z^{-4/3}. \quad \dots\dots\dots (2)$$

Thus in the steady state the form of the temperature profile is determinate from dimensional considerations.

From the similarity argument it is equally valid to carry through the dimensional analysis in terms of the local gradient $\partial\theta/\partial z$ rather than $\Delta\theta$, and the result corresponding to (1) is then

$$F \propto c \left(\frac{g}{T} \right)^{1/2} z^2 \left(-\frac{\partial\theta}{\partial z} \right)^{3/2}, \quad \dots\dots\dots (3)$$

from which of course (2) again follows as the condition for constant flux. In meteorological work it is conventional further to define the *eddy-conductivity* $K(z)$ by

$$F = -cK \frac{\partial\theta}{\partial z},$$

whence from (3)

$$K \propto z^2 \sqrt{-\frac{g}{T} \left(\frac{\partial T}{\partial z} + \Gamma \right)}, \quad \dots\dots\dots (4)$$

with Γ denoting the adiabatic lapse rate. This relation is the same as that derived by Taylor and by Sutton (1948) with L , defined by these authors as the length over which the convection currents preserve their identity, replaced by z outside the radix. It seems clear from dimensional considerations and the similarity result that L must be proportional to z , since no other length can be constructed from the independent variables of the problem. In the remainder of the paper a detailed mechanistic theory will be put forward to confirm this proportionality, to provide a physical derivation of the results obtained in this section, and a basis for discussion of the considerations which govern the constants of proportionality therein.

III. MECHANISTIC THEORY

(a) Basis of the Theory

An entirely independent line of approach may be taken by considering the detailed mechanics of the individual element of fluid which is the agent of the heat transfer. This element is to be regarded not as the finest-grained "particle" of the fluid medium, but as an aggregate of particles which have a measure of buoyancy and of vertical motion in common; the latter properties of the element will accordingly be subject to a continuous mixing with those of the environment by turbulence on a scale which is smaller than that of the element itself. This sub-elemental turbulence may take its origin in part from horizontal wind shear

(a mean horizontal flow is allowed in the present treatment provided it is not so strong that forced convection becomes dominant) but mainly as degraded motions from the buoyant motions themselves. The strength of the turbulence, being controlled by the *aggregate* of larger motions, will be a field quantity rather than one which pertains to the individual element.

On this basis it has been shown elsewhere (Priestley 1953) that the equations for the rate of change of temperature T and vertical velocity w of a buoyant element of constant size may be taken as

$$\left. \begin{aligned} \dot{w} &= \frac{g}{T_e} T' - k_1 w, \\ \dot{T} &= -w\Gamma - k_2 T', \end{aligned} \right\} \dots\dots\dots (5a)$$

$$\dots\dots\dots (5b)$$

where T' denotes temperature excess of the element over the environmental temperature T_e and dots denote time derivatives following the mean motion of the particles composing the element at a given instant. k_1 and k_2 are the mixing rates for momentum and sensible heat respectively between the element and its surroundings, proportional to the corresponding sub-elemental turbulent interchange coefficients divided by R^2 , where R characterizes the size of the element. k_1 and k_2 are to be regarded as decreasing with increasing size, but as independent of the other properties of the individual element, and as constants following a given element of constant size. These concepts have been discussed more fully in the earlier paper (*loc. cit.*). It will also be taken that $k_1 = k_2$; this step is made purely for simplification, and the analysis does not depend on it in any fundamental way. A single mixing rate k is accordingly used to identify the size of the particular element under consideration.

The equations (5) may then be differentiated following the element and T' eliminated therefrom, which leads (*loc. cit.*) to a single equation of motion

$$\ddot{w} + 2k\dot{w} + \left[\frac{g}{T_e} \left(\frac{\partial T_e}{\partial z} + \Gamma \right) + k^2 \right] w = 0, \dots\dots\dots (6)$$

provided that T'/T_e is small, which is always the case under natural conditions. The temperature of the element may then be obtained from (5a) in the form

$$T' = \frac{T_e}{g} (\dot{w} + kw). \dots\dots\dots (5a)$$

It is required to solve these equations for a form of environmental temperature profile which is typical of conditions above a heated surface, and thence to formulate the heat flux by recombining the different solutions for elements of different sizes moving through the reference level to which the heat flux relates. The profile of temperature or of potential temperature θ will be taken to be of the form

$$\frac{g}{\theta} \frac{\partial \theta}{\partial z} = \frac{g}{T_e} \left(\frac{\partial T_e}{\partial z} + \Gamma \right) = -Cz^{-\delta}, \dots\dots\dots (7)$$

where δ and C are constants characterizing its shape and the general magnitude of the gradient respectively. Since the absolute θ in the denominator varies relatively slowly with height, this is in effect equivalent to the general form deduced in Section II, but to preserve independence of the two treatments (7) will here be regarded as an assumed form with sufficient flexibility to cover all practical cases to which the theory will be applied. C is clearly positive, but no theory has previously provided a value of δ . From the dimensional analysis a value of $4/3$ is suggested as of special significance and this is confirmed in the following, but for the present δ will be allowed to take any positive value.

(b) *Solution of the Equations*

Equation (6) may now be written

$$\ddot{z} = Cz^{-\delta} \dot{z} - k^2 \dot{z} - 2k\ddot{z}. \quad \dots\dots\dots (8)$$

We shall consider here only the ascending elements; the principal results are modified only by a numerical multiplier when the descending air is allowed for (Section VI (a)). The effects of any externally produced irregularities will be disregarded. That is to say, an individual element will be treated as starting from rest at some level z_0 with a temperature equal to that of the environment at that level. The consequences of the removal of this last assumption are discussed in Section VI (b). From (5a) it follows that $\ddot{z}=0$ at z_0 , whence (8) can be integrated to

$$\dot{z} = w = \frac{C}{1-\delta} (z^{1-\delta} - z_0^{1-\delta}) - k^2(z - z_0) - 2k\dot{z}. \quad \dots\dots\dots (9)$$

This is now integrated with respect to z from the starting level z_0 to the reference level z_1 at which the heat flux is required, and we obtain after some rearrangement

$$\frac{1}{2}w^2 = Cz_1^{2-\delta} \left\{ \frac{1 - \left(\frac{z_0}{z_1}\right)^{1-\delta}}{1-\delta} - \frac{1 - \left(\frac{z_0}{z_1}\right)^{2-\delta}}{2-\delta} \right\} - \frac{1}{2}k^2z_1^2 \left(1 - \frac{z_0}{z_1}\right)^2 - 2k \int_{z_0}^{z_1} w dz. \quad \dots\dots\dots (10)$$

This is the equation* for w , regarded as a function of z_1 and dependent also on C , δ , k , and z_0 , in a form which lends itself to numerical solution by successive approximation when the values of the independent quantities are specified.

It is possible, without solving explicitly, to infer the form of dependence of w on certain of these independent variables. For this purpose we shall define a critical mixing rate k_c by

$$k_c^2 = Cz_1^{-\delta}. \quad \dots\dots\dots (11)$$

* The equation is not valid for $\delta=1$ or $\delta=2$, when logarithmic terms would appear; the alternative forms are not set out here, but they have been used whenever applicable in all calculations whose results are quoted hereafter.

This rate has local physical significance in that elements for which $k < k_c$, that is, which are larger than the critical size to which k_c refers, are absolutely buoyant or unstable at the level z_1 (Priestley, loc. cit.). The term containing k^2 in (10) may now be expressed as

$$-\frac{1}{2}Cz_1^{2-\delta}\left(\frac{k}{k_c}\right)^2\left(1-\frac{z_0}{z_1}\right)^2,$$

and k in the last term as $(k/k_c)C^{1/2}z_1^{-\delta/2}$. It may then be seen on close scrutiny that w will take the functional form

$$w = C^{\frac{1}{2}}z_1^{\frac{2-\delta}{2}}n_1, \dots\dots\dots (12)$$

where n_1 is a dimensionless number which depends only on δ and the values of z_0/z_1 and k/k_c , and can be calculated from assigned values of these ratios. The formal demonstration that there is a solution of this form is obtained by substitution from (12) in (10) which then may be reduced to an equation for n_1 in the form

$$\frac{1}{2}n_1^2 = \frac{1 - \left(\frac{z_0}{z_1}\right)^{1-\delta}}{1-\delta} - \frac{1 - \left(\frac{z_0}{z_1}\right)^{2-\delta}}{2-\delta} - \frac{1}{2}\left(\frac{k}{k_c}\right)^2\left(1-\frac{z_0}{z_1}\right)^2 - 2\frac{k}{k_c}\int_{z=z_0}^{z=z_1} n_1\left(\frac{z}{z_1}\right)^{\frac{2-\delta}{2}}d\left(\frac{z}{z_1}\right), \dots\dots\dots (13)$$

which shows clearly the quantities on which n_1 must depend. The uniqueness of the solution may be established from (9), for, if w_1 and w_2 are two solutions,

$$\dot{w}_1 - \dot{w}_2 = -2k(w_1 - w_2),$$

or

$$w_1 - w_2 = \text{const. } e^{-2kt},$$

and, since both satisfy $w=0$ at $t=0$, $w_1=w_2$.

Turning now to the temperature excess of the element, it is seen from (5a) and (9) that

$$T' = \frac{T_e}{g} \left\{ Cz_1^{1-\delta} \left(\frac{1 - \left(\frac{z_0}{z_1}\right)^{1-\delta}}{1-\delta} \right) - k^2 z_1 \left(1 - \frac{z_0}{z_1} \right) - kw \right\}, \dots\dots (14)$$

which can be calculated after w has been obtained from (10). From similar reasoning to the above,

$$T' = \frac{T_e}{g} Cz_1^{1-\delta} n_2, \dots\dots\dots (15)$$

whence

$$wT' = \frac{T_e}{g} C^{\frac{3}{2}} z_1^{\frac{4-3\delta}{2}} n_3, \dots\dots\dots (16)$$

where n_2 and n_3 , like n_1 , are numbers dependent only on δ and on z_0/z_1 and k/k_c .

(c) *Formulation of the Heat Flux and other Field Statistics*

To formulate the heat flux F equation (16) must be averaged over all particles crossing the reference plane z_1 at a given instant. The elements to which the particles belong have a variety of sizes, characterized by k , and of starting levels z_0 , so that a double averaging process is involved. We introduce a probability function $f(k, z_1, C, \delta)$ to represent the probability that a particle at z_1 will, for a given C and δ , belong to an element of size k ; the probability distribution is in effect a frequency distribution for k weighted according to element size. The four arguments of f permit of only two dimensionless combinations, δ and $kC^{-1/2}z_1\delta^{1/2}$ or k/k_c , so that the probability distribution can be represented by $f(k/k_c, \delta)$, where

$$\int_0^\infty f\left(\frac{k}{k_c}, \delta\right) d\left(\frac{k}{k_c}\right) = 1.$$

The averaging of (16), for a given value of δ , is now carried out first over all admissible values of the starting level z_0 for a fixed value of k ; this step will be indicated by a bar (a bar will also be used to denote averages taken over all particles, but there should be no confusion). The average is then taken over all values of k , and the result is seen to be

$$\frac{F}{\rho c_p} = \overline{wT'} = \frac{T_e}{g} C^{3/2} z_1^{\frac{4-3\delta}{2}} N_3. \quad \dots\dots\dots (17)$$

N_3 is another number, which is given in terms of n_3 by

$$N_3 = \int_0^\infty f\left(\frac{k}{k_c}, \delta\right) \overline{n_3\left(\frac{k}{k_c}, \delta, \frac{z_0}{z_1}\right)} d\left(\frac{k}{k_c}\right).$$

To understand the meaning of the term "admissible value" it is necessary to go more closely into the details of the motion and temperature behaviour. This involves numerical calculation in the equations set out above; some of the computations are presented in another context in Section V, but here it will be sufficient to state the main features and illustrate them by a diagram (Fig. 1).

Consider, first, elements all of a given size; that is, fix the value of k , which will be the critical value (equation (11)) for some level P on the profile. For an element at rest at P the tendency to mix will just compensate for the instability of the lapse rate, and the element will just not move. Elements at rest above P are in a less unstable environment and will not move spontaneously. At Q , a short distance below P , the instability at first outweighs the mixing and an element here will move upwards, reaching the level of P with finite velocity and temperature excess; from here onwards the coefficient of w in (6) has changed sign and becomes increasingly positive, the mixing will outweigh the decreasing environmental instability, and the element will be brought to rest at some higher level Q' . The final approach to rest is of the exponential type and \dot{w} and hence T' tend to zero simultaneously, the element being completely merged into its new surroundings. An element from a lower level R will reach the level of P with a larger w and T' than that from Q and so will penetrate above Q' , to R' ;

and so on, SS' , TT' , etc., representing paths of elements starting from successively lower levels.

Now consider all elements as they cross the reference level z_1 . If $k < k_c$, P will lie above z_1 , and an element of this size reaching z_1 may have started at any level below z_1 . For these values of k , therefore, the averaging of n_3 in the integral for N_3 must extend from $z_0 = z_1$ down to $z_0 = \Delta$. If, however, $k > k_c$, P will lie below z_1 as in the figure, and it is evident that elements reaching z_1 must have started from some level considerably below z_1 . In Figure 1, an

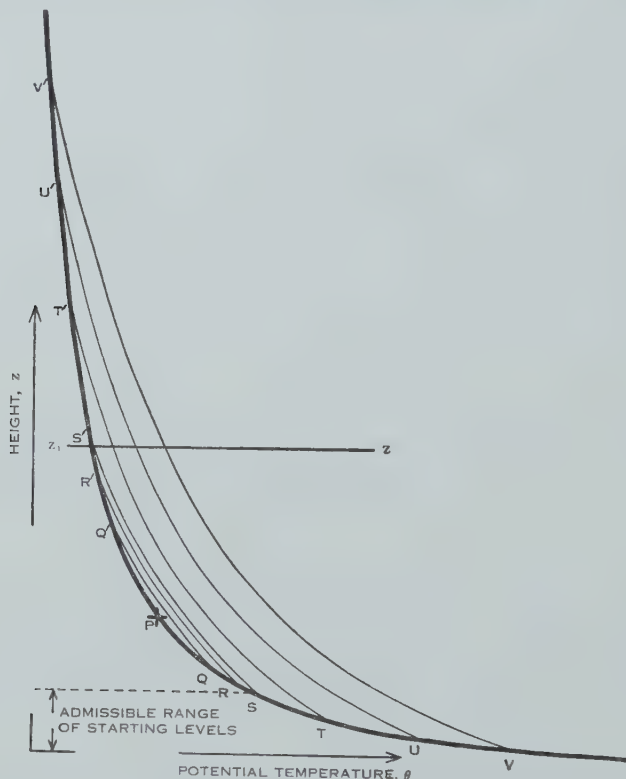


Fig. 1.—Schematic representation of paths of elements. These are represented as thin lines against the thick line for the mean profile of the environment.

element starting from between z_1 and S would fail to reach z_1 . In this case the averaging of n_3 over z_0 must extend only over the "admissible range" from the level S down to $z_0 = \Delta$. The level S will depend on k , and the value of z_0 appropriate to it may be derived most readily from (9) or (14). It is given by

$$\left(\frac{k}{k_c}\right)^2 = \frac{1 - \left(\frac{z_0}{z_1}\right)^{1-\delta}}{(1-\delta)\left(1 - \frac{z_0}{z_1}\right)},$$

which shows that z_0/z_1 , at the top of the admissible range, is determined by δ and k/k_c only.

It has now been shown that the upper limit of the range of values of z_0 , over which n_3 must be averaged, is either equal to z_1 (for $k < k_c$) or (for $k > k_c$) to a given submultiple of z_1 which depends only on k/k_c and δ . The lower limit is $z_0 = \Delta$, and this is taken to be effectively equivalent to $z_0 = 0$, assuming convergence of the integral as $z_0 \rightarrow 0$ which will be established later in the paper (Section V). It now follows from the result established above that \bar{n}_3 will be a function of k/k_c and δ only; that is, *it will not otherwise contain z_1* . It then follows further that N_3 , which is essentially a weighted average of \bar{n}_3 with respect to k/k_c , is independent of z_1 .

It has thereby been proved that N_3 , though dependent on δ and the form of the unknown function f , is independent of z_1 and of C ; and hence that the indices of z_1 and C appearing in (17) correctly represent the functional dependence of heat flux on height and on lapse rate.

In a similar way, corresponding expressions may be derived for other statistical representations of the temperature and velocity distributions at the level z_1 . Those to which subsequent appeal will be made are

$$\bar{w} = C^{\frac{1}{2}} z_1^{\frac{2-\delta}{2}} N_1, \quad \dots \dots \dots (18)$$

$$\bar{T}' = \frac{T_e}{g} C z_1^{1-\delta} N_2, \quad \dots \dots \dots (19)$$

$$\sigma_w = \sqrt{\bar{w}^2} = C^{\frac{1}{2}} z_1^{\frac{2-\delta}{2}} N_4, \quad \dots \dots \dots (20)$$

$$\sigma_{T'} = \sqrt{\bar{T}'^2} = \frac{T_e}{g} C z_1^{1-\delta} N_5, \quad \dots \dots \dots (21)$$

where in general

$$N_i = \int_0^\infty f\left(\frac{k}{k_c}, \delta\right) \bar{n}_i d\left(\frac{k}{k_c}\right), \quad \dots \dots \dots (22)$$

and the derivation of n_4 and n_5 from (10) and (14) is along similar lines to that of n_1 , n_2 , and n_3 .

N_1 to N_5 have the common property that they depend on δ and on the form of f , but not on z_1 and C . The foregoing analysis has thus separated into two parts the problem of describing the fields of temperature and motion. The first part relates to the shape of the temperature profile, that is, the value of δ , and to those relationships which depend on it alone, namely the form of F , σ_T , etc., as functions of height and of lapse rate. This part will be treated in Sections III (e) and IV, and the second, relating to the absolute values of the multiplying factors N_1 to N_5 which depend on f as well as on δ , in Section V.

In passing, it has been established from equations (10) and (14) that z_1 is itself the length scale which characterizes all aspects of the temperature and velocity fields at the level z_1 ; in particular that the admissible range of initial levels is proportional to z_1 . It follows that L , the convective length envisaged

by Taylor and Sutton, is proportional to z_1 , and that the form of the relation (4) proposed here can be reconciled with theirs. Up to this point there is no conflict between the two treatments, but a relation introduced by Sutton (1948) at a later stage, that $L \propto z^{1.35}$, is not in harmony with the theory offered here.

(d) *Forced Convection and Radiation*

The relations (17)–(21) are derived on the basis that the statistics of the temperature and vertical velocity distribution are determined from the operation of free convection alone. In nature, both radiation and forced convection exert complicating influences. Whereas the effects of the former lie clearly outside the bounds of the present treatment, this is not entirely so in the case of forced convection; rather in the atmospheric problem is it idle to attempt too fine a distinction between free and forced convection, since they occur together and their effects are inseparable. The extent to which the vertical movement in an unstably stratified boundary layer exceeds that in a neutral layer at the same wind strength would appear to be a consequence solely of the processes which are here under consideration. For this reason it might be expected that the predictions of the theory would have approximate validity when a wind is blowing, provided it is not too strong or the lapse rate too weak.

In short, the results should extend to finite values of the Richardson number, but at the present stage it is probable that experiment will be more successful than theory in indicating the limits of validity. The relations (18) and (20) will evidently cease to hold when the buoyant motions no longer outweigh those deriving from wind shear but it may well be that at this stage the temperature fluctuations, and hence the heat flux, will still be dominated by the former. Thus the range of applicability of the other predictions may extend considerably beyond that of (18) and (20).

(e) *Specification of the Index δ*

From (17) it follows immediately that F is constant with height when $\delta=4/3$, and that it decreases and increases with height when δ is greater and less than $4/3$ respectively. Thus, in the absence of radiation influences and strong forced convection, $\delta=4/3$ will characterize the steady state, $\delta>4/3$ a condition in which the air is warming, and $\delta<4/3$ a condition in which it is cooling.

In practice, in the meteorological case at any rate, an upwards flux of any considerable magnitude will only vary with height very slowly. For a modest F of about 10 mW/cm^2 , a 5 per cent. change in flux between 1 and 2 m would represent heating or cooling at 15°C/hr , a rate which is never approached; yet such changes would be represented by $\delta=1.29$ and $\delta=1.38$. The conclusion is that δ will in practice be constrained to a value quite close to $4/3$. This value will accordingly be adopted as a definite prediction to be tested, together with the consequences which follow from it in equations (17)–(21), against actual measurements of the profile and of F , σ_T , and σ_w .

An exception to the value $4/3$ may be expected from the bottom layer (below about 1 m), which normally under lapse conditions is strongly warmed by radiation (Robinson 1950). To maintain the quasi-steady state in this layer the

effects of radiation and convection must approximately compensate, and a δ lower than $4/3$ is therefore to be expected.

It is of interest to remark that an index of $4/3$ and approximately constant flux imply that the eddy-conductivity K is proportional to $z_1^{4/3}$, which resembles theoretical and empirical results in other problems of turbulent transfer over a very wide range of scales.

IV. EXPERIMENTAL DATA

(a) *The Mean Temperature Profile*

The first prediction to be tested is the temperature profile law, $\delta=4/3$. For a few clear days in summer at Leafield, England, Sutton (1948) finds $\delta=1.75$ but an analysis by Deacon (1948) of a far larger number of observations from the same site yields values ranging from 1.23 to 1.53 at different ranges of wind speed. Deacon (unpublished data) has analysed data from other sites with the results shown in Table 1.

TABLE 1
ANALYSIS OF TEMPERATURE PROFILES BY DEACON

Location	Height Range (m)	Reference for Original Data	Conditions	δ
Porton, England	1.2-17	Deacon (1953)	256 clear days (May, June, July) 1923-43, 1100-1400 hr	1.35
Ismailia	1.1-46	Flower (1937)	All days Apr.-Sept. 1932, 1400-1500 hr	1.43
Manor, Texas	2-25	Gerhardt <i>et al.</i> (1948), Gerhardt (1949)	Fair or fine days,* July-Oct. 1948, 1100-1300 hr	1.25
Edithvale, Victoria	1-30	Unpublished	Fine days, Nov. 1951-Mar. 1952, 1200-1600 hr	1.25

* With wind at 12 m less than 4.5 m/sec.

There is some scatter, but much of this may result from attributing too fine an accuracy to the experimental measurements. Extraction of a value of δ from a set of measured temperatures involves essentially the evaluation of second differences between temperatures which are subject to errors of calibration, retention of calibration, standardization of shielding and of aspiration; such errors may be to a considerable extent systematic over the period of observation. The point may be illustrated by Table 2 in which are given two calculated profiles for $\delta=1.5$ and 1.3. These two profiles, which represent a strong lapse rate, differ nowhere by more than 0.11 °F, a measure of accuracy which can hardly be claimed for much of the original data.

A separate analysis has been made of some published data on strong lapse profiles in the layers mainly below 1 or 2 m where radiational heating becomes large (Robinson, *loc. cit.*) and the present theory would predict a value of δ less than $4/3$. The results are shown in Table 3. Save in the case of the Porton

data, the number of levels of observation were more than the three required to provide a value of δ , and the value given was obtained by least squares applied to the logarithms of the heights and measured temperatures.

The data from Poona are of particular interest because they provide a simultaneous and detailed profile both above and below the level where radiation effects become important. Analysis of the profile in two overlapping regions indicates a considerable difference in δ .

TABLE 2
CALCULATED TEMPERATURE PROFILES FOR $\delta=1.5$ AND 1.3

Height (m)	2	4	8	16	32
Temperature ($^{\circ}\text{F}$), $\delta=1.5$..	73.45	72.92	72.16	71.06	69.41
„ ($^{\circ}\text{F}$), $\delta=1.3$..	73.55	72.89	72.05	70.96	69.51

The individual mean profiles summarized in Tables 1 and 3 are shown diagrammatically* in Figure 2. A striking feature is that all the gradients from the stations in the lower latitudes (Ismailia, Manor, Edithvale, and Poona) lie approximately on *the same* profile above 1 m, and a single dashed line of slope $\delta=4/3$ has been drawn by eye to illustrate the closeness of fit of the prediction.

TABLE 3
PROFILES AT LOW LEVELS

Location	Height Range (cm)	Reference	Conditions	δ
Porton, England	2.5-120	Best (1935, Table VI)	58 clear occasions in June, 1100-1300 G.M.T.	1.22
Cambridge, England	25-200	Pasquill (1949)	15 lapse profiles	1.05
Kew, England	5-202	Rider & Robinson (1951)	15 lapse profiles	1.40
Poona, India	1-180	Ramdas (1953)	April, hr of max. temp.	0.93
„ „	90-750	„ „	„ „ „	1.29

The full lines I, II, and III represent the calculated regression lines for the Poona, Cambridge, and Kew data respectively, the first being shown in two parts relating to the two overlapping ranges.

On the whole the theoretical predictions, that δ shall be $4/3$ above about 1 m and less below, fit the facts within the expected tolerance.

* Where the original data are given as temperatures at, or differences between, fixed levels, the slope of the chord has been plotted against the geometric mean height. This is strictly legitimate only for $\delta=1$, but the error involved in other cases is very small.

(b) Heat Flux and Temperature and Velocity Fluctuations

With substitution of the value $\delta=4/3$ in (17)–(21) the variations of the flow and temperature characteristics with height, as well as with lapse rate, become determinate.

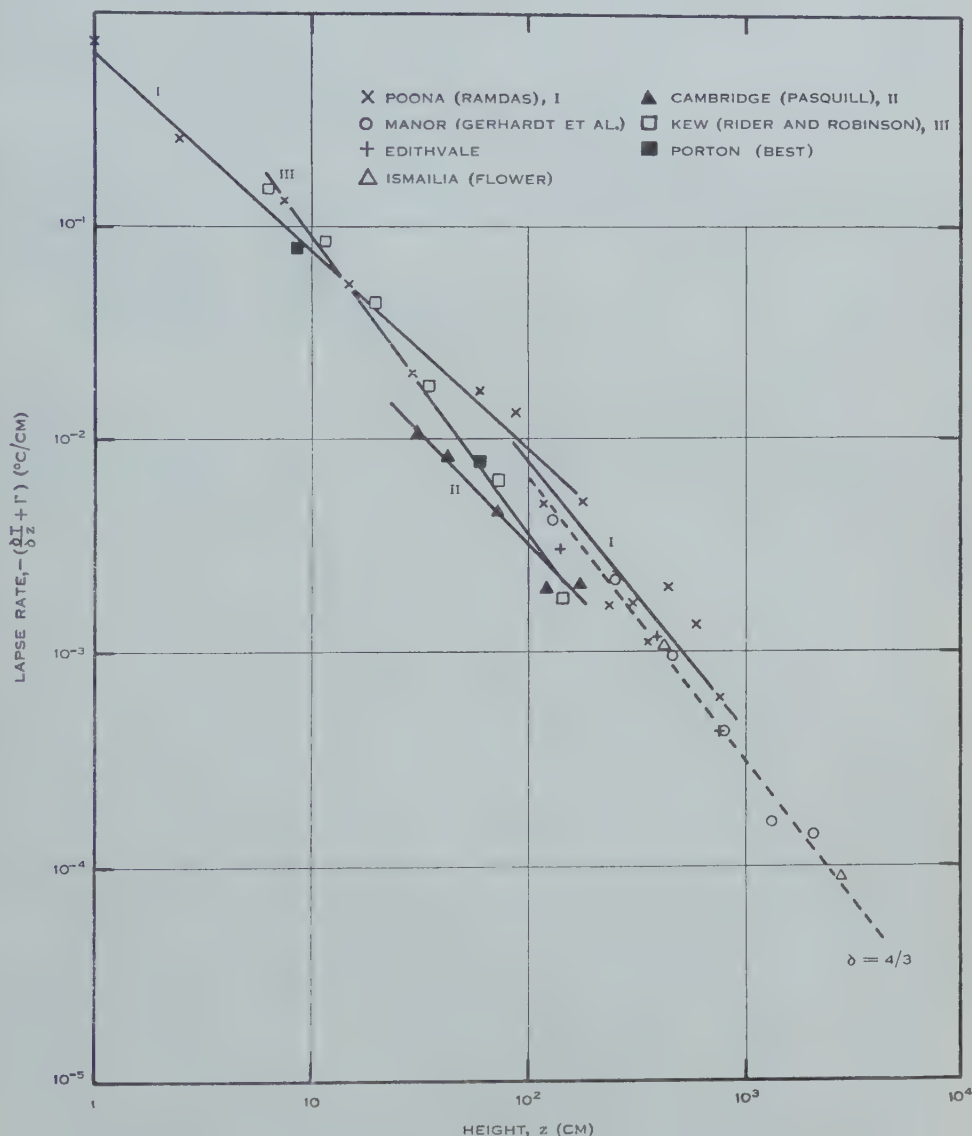


Fig. 2.—Lapse profiles from 1 cm to 100 m.

The published literature does not contain the information on which a satisfactory test of these theoretical predictions can be made. However, a group of workers in the Section of Meteorological Physics, C.S.I.R.O., has made measurements of the quantities involved at Edithvale, Vic., using the instruments

TABLE 4
EXPERIMENTAL DATA FROM EDITHVALE ON 3.iii.1952

Run No.	1	2	3	4	5	6	7	8	9	10	11	12
Wind velocity at 2 m (m/sec) ..	4.1	4.2	4.5	3.9	4.5	4.9	4.6	5.1	4.4	4.6	4.4	4.4
Potential temperature gradient — $\partial\theta/\partial z$ at 1.5 m ($^{\circ}\text{C}/\text{m}$)	0.83	0.71	0.71	0.59	0.55	0.58	0.51	0.50	0.24	0.22	0.18	0.13
Height of exposure of fine structure apparatus (m)	1.92	29	1.92	29	1.92	29	1.92	29	1.92	29	1.92	29
Heat flux F (mW/cm ²) ..	19.9	17.0	20.1	23.4	14.4	37.7	10.9	4.8	5.4	9.9	2.8	0.4
σ_T ($^{\circ}\text{C}$)	1.00	0.38	0.88	0.32	0.70	0.47	0.56	0.18	0.27	0.14	0.20	0.11
σ_w (cm/sec)	39	70	41	92	36	91	39	58	39	84	37	55
Negative Richardson number at 1.5 m	0.14	0.11	0.11	0.12	0.10	0.07	0.06	0.04	0.04	0.02	0.02	0.02

and techniques developed by Swinbank (1950). Although at the time of writing much of the data still awaits analysis, sufficient is available to allow preliminary tests of the predictions.

In the experiments the apparatus designed by Swinbank for the measurement, among other things, of heat flux was exposed for 5-min periods usually, though not always, alternately at two different heights above ground level. From the records obtained, the values of F , σ_T , and σ_w have been extracted. Auxiliary observations included the wind and temperature profile; for purposes of the present tests these have been reduced by an objective method to temperature gradient and Richardson number at a standard level (1.5 m).

In Table 4 are presented the results of a sequence of measurements made on a single day in which a large range of lapse rates was experienced, conditions otherwise remaining steady.

To analyse these results, the logarithms of the measured quantities F , σ_T , and σ_w have been taken and the regression line

$$\log \lambda + \mu \log C + \nu \log z$$

fitted to them, the values of μ and ν and of their standard errors being calculated by the method of least squares. The regressions are given in Table 5, standard errors being shown in brackets and the theoretical relations set alongside.

TABLE 5
TEST OF THEORETICAL PREDICTIONS

Theoretical Relation	Empirical Result
$F \propto C^{3/2}$, independent of z	$F \propto C^{1.73(\pm 0.32)} z^{0.00(\pm 0.14)}$
$\sigma_T \propto C z^{-1/3}$	$\sigma_T \propto C^{0.88(\pm 0.10)} z^{-0.26(\pm 0.04)}$
$\sigma_w \propto C^{1/2} z^{1/3}$	$\sigma_w \propto C^{0.02(\pm 0.06)} z^{0.24(\pm 0.03)}$

Only the σ_w relation departs radically from its predicted form, and this was expected since a wind of 4–5 m/sec blew steadily throughout, while the theoretical σ_w relates only to those motions attributed to buoyancy. The results lend support to the speculations of Section III (d) that in fair winds the buoyant motions may constitute only a minor part of the vertical motion, but that it is this minor part which is correlated with the temperature fluctuations and so dominates the heat flux.

The agreement for F and σ_T justifies a more severe test by using less homogeneous material. All the complete sets of measurements so far available from other occasions are set out in Table 6, the only criterion for acceptance being that $|Ri|$ should be not less than 0.02.

The regression relations, from the whole set of 30 runs, are

$$F \propto C^{1.68(\pm 0.20)} z_1^{-0.04(\pm 0.08)},$$

$$\sigma_T \propto C^{0.92(\pm 0.09)} z_1^{-0.25(\pm 0.04)}.$$

TABLE 6
EXPERIMENTAL DATA FROM OCCASIONS OTHER THAN 3.iii.52

Date	21.i.52					30.i.52			18.iii.52			4.iii.52		20.ii.53			25.ii.52	
Run No. (for identification)	1	2	3	4	5	1	2	3	3	5	6	47	48	4	5	9	7	38
Wind at 2 m (m/sec)	5.24	5.12	5.54	5.26	5.31	3.10	3.41	3.12	1.86	1.90	1.91	4.39	4.10	1.56	1.87	2.01	4.45	5.08
$-\partial\theta/\partial z$ at 1.5 m ($^{\circ}\text{C}/\text{m}$)	0.71	0.72	0.65	0.63	0.60	0.63	0.59	0.57	0.24	0.20	0.25	0.25	0.22	0.16	0.38	0.42	0.51	0.33
Height z (m)	1.5	29	1.5	29	1.5	1.5	29	1.5	1.5	1.5	23	1.92	29	1.5	1.5	1.5	1.85	1.85
F (mW/cm ²)	18.7	21.0	22.3	17.1	23.3	20.4	6.4	21.1	4.4	5.3	2.2	7.0	5.1	0.5	2.6	6.8	16.3	11.3
σ_T ($^{\circ}\text{C}$)	0.80	0.435	0.71	0.345	0.885	0.87	0.485	0.81	0.395	0.470	0.187	0.345	0.126	0.135	0.208	0.526	0.82	0.40
$-R_z$ at 1.5 m	0.04	0.05	0.04	0.03	0.03	0.18	0.11	0.12	0.24	0.11	0.24	0.04	0.06	0.20	0.22	0.33	0.02	0.02

The agreement is on the whole a little better than in Table 5, and the uncertainty of the index of C in F is much reduced. It is desirable that more measurements be made, sufficient to be analysed in ranges of the Richardson number, before the goodness of fit be finally judged.

Measurements of σ_T from 7 to 75 m at Leafield by Johnson and Heywood (1938) have been analysed by Sutton (1948) who finds $\sigma_T \propto z^{-0.4}$, but this result must carry less weight owing to the much lower sensitivity of the instruments used in the early measurements.

It is usual to regard F as determined by some suitable combination of the characteristics of the wind and temperature profiles, and to this end the data of Tables 4 and 6 may be analysed in another way, by calculating the partial correlations of F with each of the three characteristics C , Ri , and u (wind speed). The results are

$$r_{FC,uRi} = +0.63(\pm 0.11 \text{ S.E.}),$$

$$r_{Fu,C Ri} = +0.11(\pm 0.18),$$

$$r_{FRi,Cu} = -0.02(\pm 0.18).$$

The total correlation between F and C was $+0.77$. For two quantities between which the true relation is non-linear, this is a very high coefficient. And the general thesis of this paper, that in the more unstable situations the heat flux is dominated by its dependence on lapse rate alone, appears to be established under the conditions of these experiments.

V. THE CONSTANTS OF PROPORTIONALITY

(a) *Upper Limits to the Constants*

It remains to examine in greater detail the constants of proportionality, N_1 to N_5 , which occur in equations (17)–(21). Although the exact values of these numbers have been shown to depend on the form of the function f , and therefore must remain unknown from theoretical considerations until this function is provided, useful inferences can be made about their relative magnitudes and it will be shown that their orders of magnitude also are determinable from the theory at its present stage.

In continuing the theoretical treatment of Sections III (b) and III (c) we shall retain the freedom there allowed to the value of δ , delaying the identification $\delta=4/3$ as late as possible.

An element starting from z_0 will have a velocity and temperature excess at z_1 smaller than if there were no mixing ($k=0$). We shall denote the values attained on the latter basis by \bar{w}^* and \bar{T}^* , and refer to them as the *free* velocity and temperature excess. The free values, when substituted in the expressions for F , σ_T , etc. provide an upper limit to the values of these quantities. The device is equivalent to the supposition of a special form for the function $f(k/k_c, \delta)$, assigning f the value 1 for $k=0$ and $f=0$ for all other k . We shall evaluate this upper limit and, by comparing it with measured values, obtain a quantitative assessment of the importance of the mixing process.

From (10) and (14), then

$$\frac{1}{2}\dot{w}^2 = Cz_1^{2-\delta} \left\{ \frac{1 - \left(\frac{z_0}{z_1}\right)^{1-\delta}}{1-\delta} - \frac{1 - \left(\frac{z_0}{z_1}\right)^{2-\delta}}{2-\delta} \right\}, \quad \dots\dots (23)$$

$$\dot{T}' = \frac{T}{g} Cz_1^{1-\delta} \left\{ \frac{1 - \left(\frac{z_0}{z_1}\right)^{1-\delta}}{1-\delta} \right\}, \quad \dots\dots\dots (24)$$

and the substitution of these values leads to relations identical with (17)–(21) with the numbers N_1 to N_5 replaced by integrals I_1 to I_5 given by

$$I_1(\delta) = \frac{z_1}{z_1 - \Delta} \int_{\Delta/z_1}^1 \sqrt{2} \left\{ \frac{1 - \theta^{1-\delta}}{1-\delta} - \frac{1 - \theta^{2-\delta}}{2-\delta} \right\}^{\frac{1}{2}} d\theta,$$

$$I_3(\delta) = \frac{z_1}{z_1 - \Delta} \int_{\Delta/z_1}^1 \sqrt{2} \left\{ \frac{1 - \theta^{1-\delta}}{1-\delta} - \frac{1 - \theta^{2-\delta}}{2-\delta} \right\}^{\frac{1}{2}} \left\{ \frac{1 - \theta^{1-\delta}}{1-\delta} \right\} d\theta,$$

and I_2, I_4 , and I_5 in similar form. θ has been written for z_0/z_1 (z_1 fixed while z_0 varies) and Δ is defined in Section I. For $z_1 \gg \Delta$, so long as the integrals converge as the lower limit tends to zero, we may in practice take it as zero and drop the factor $z_1/(z_1 - \Delta)$ outside the integral sign.

These integrals have been evaluated, the curve for $I_3(\delta)$ being shown in Figure 3 and values of I_1 to I_5 at discrete values of δ in Table 7.

TABLE 7
UPPER LIMITS TO VALUES OF N_1 TO N_5

δ	0.7	0.9	1.1	4/3	1.5	5/3
I_1	0.67	0.74	0.81	0.94	1.02	
I_2	0.77	0.91	1.11	1.5	2.0	
I_3	0.82	1.22	1.91	4.0	8.4	∞
I_4	0.875	0.93	1.07	1.34	1.63	
I_5	0.98	1.23	1.67	3.0	∞	

None of the integrals diverge for $\delta=4/3$, the case of greatest physical interest. This establishes the required convergence property for \bar{n}_i , which was referred to in Section III (c), for $k=0$. For $k>0$, the convergence holds *a fortiori*.

Reverting now and hereafter to the special case $\delta=4/3$, it has been shown that upper limits to the statistical quantities are provided by the theory as

$$\frac{\dot{w}}{2} = 0.94 C^{1/2} z_1^{1/3}, \quad \dots\dots\dots (25)$$

$$\dot{T}' = \frac{3}{2} \frac{T}{g} C z_1^{-1/3}, \quad \dots\dots\dots (26)$$

$$\dot{F} = 4\rho c_p \frac{T}{g} C^{3/2}, \quad \dots\dots\dots (27)$$

$$\dot{\sigma}_w = 1.34 C^{1/2} z_1^{1/3}, \quad \dots\dots\dots (28)$$

$$\dot{\sigma}_T = \frac{3T}{g} C z_1^{-1/3}. \quad \dots\dots\dots (29)$$

(b) *Correction Factors to the Upper Limits*

Comparing (27) and (29) with the data in Tables 4 and 6, we find

$$\text{Geometric mean ratio } \frac{\bar{F}^*}{\bar{F}(\text{observed})} = 4.7,$$

$$\text{Geometric mean ratio } \frac{\sigma_T^*}{\sigma_T(\text{observed})} = 3.1.$$

That the upper limits of \bar{F} and σ_T are not divorced by an order of magnitude from the measured values is an achievement, considering that the estimate in

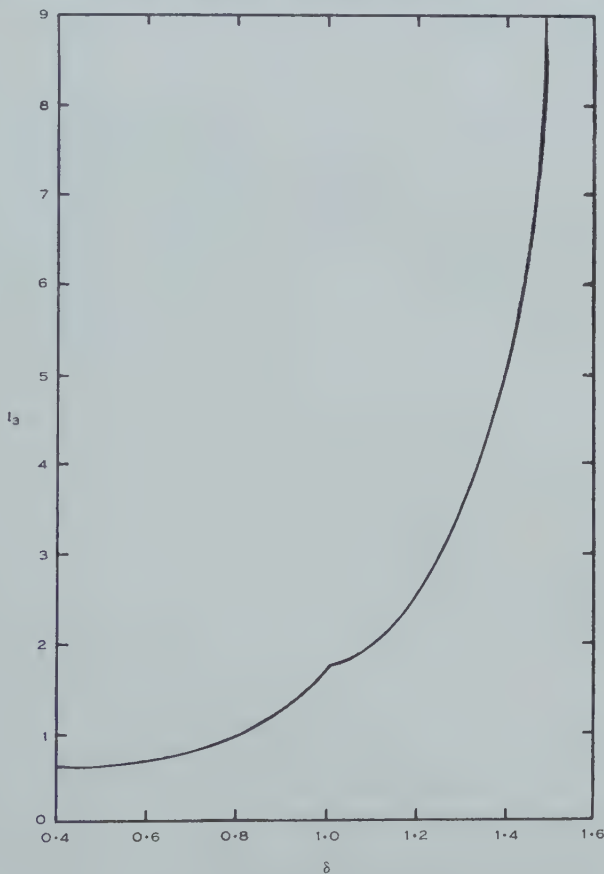


Fig. 3.—Values of flux integral (I_3) with no mixing.

effect invokes the measured lapse rate only. Further examination of the “free” motion device is suggested, in order to see how much can be stated about the correction factors to \bar{F}^* etc., when mixing is allowed for.

Following (12) we may write, for $\delta=4/3$,

$$w = C^{1/2} z_1^{1/3} n_1 \left(\frac{k}{k_c}, \frac{z_0}{z_1} \right), \quad \dots \dots \dots (30)$$

and similarly

$$w = C^{1/2} z_1^{1/3} n_1^* \left(\frac{z_0}{z_1} \right), \dots\dots\dots (31)$$

so that

$$w/\dot{w} = n_1/\dot{n}_1,$$

which will be referred to as the mixing factor for w ; numbers \dot{n}_2 to \dot{n}_5 and corresponding mixing factors may be similarly defined. All the mixing factors may be calculated by substitution from (30) etc. in (10) and (14), and from (31) etc. in (23) and (24).

The calculation has been carried out for z_0/z_1 from 1/2 to 1/200, over the full significant range of k/k_c , for w from equation (10) and for T' from (14). The range of z_0 chosen was such as to contribute three-quarters of the total value of the flux integral I_3 , most of the remaining quarter coming from $z_0 < z_1/200$, to which the calculation may be extended as required. Figures 4 and 5 show isopleths of the factors n_1/\dot{n}_1 and n_2/\dot{n}_2 , drawn on diagrams with z_0/z_1 as abscissa and k/k_c as ordinate. Table 8 gives in detail the contributions to \dot{F} from various ranges of z_0/z_1 (or θ), indicating the potential relative importance of the different possible levels of origin.

TABLE 8.
CONTRIBUTION TO \dot{F} OR I_3 FROM VARIOUS RANGES OF $\theta = z_0/z_1$

z_0/z_1 from ..	0	0.002	0.005	0.01	0.05	0.1	0.2	0.5
to	0.002	0.005	0.01	0.05	0.1	0.2	0.5	1
Percentage of integral	14.5	7	7.5	29	15	12	12	3

These calculations complete the information as far as the theory at present allows, enabling the multiplying constants in (17)–(21) to be immediately worked out once the single remaining unknown, the function $f(k/k_c)$, is assigned. An approximate simplification is possible by use of the features of Figures 4 and 5 that the mixing factors are found to depend on k/k_c much more strongly than on z_0/z_1 . That is to say, the variation of n_i with z_0/z_1 is largely already taken up in \dot{n}_i , which allows as a working approximation

$$\frac{n_i\left(\frac{k}{k_c}, \frac{z_0}{z_1}\right)}{\dot{n}_i\left(\frac{z_0}{z_1}\right)} \sim \frac{n_i\left(\frac{k}{k_c}, \frac{1}{25}\right)}{\dot{n}_i\left(\frac{1}{25}\right)},$$

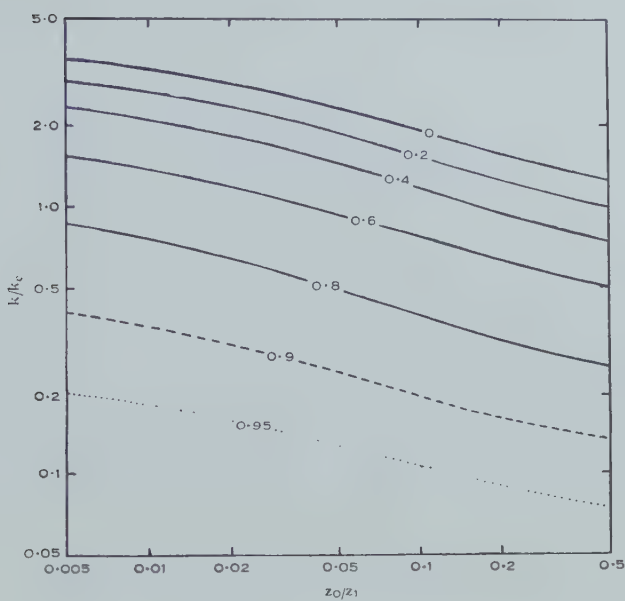


Fig. 4.—Mixing factor for velocity ($w/w = n_1^*/n_1$).

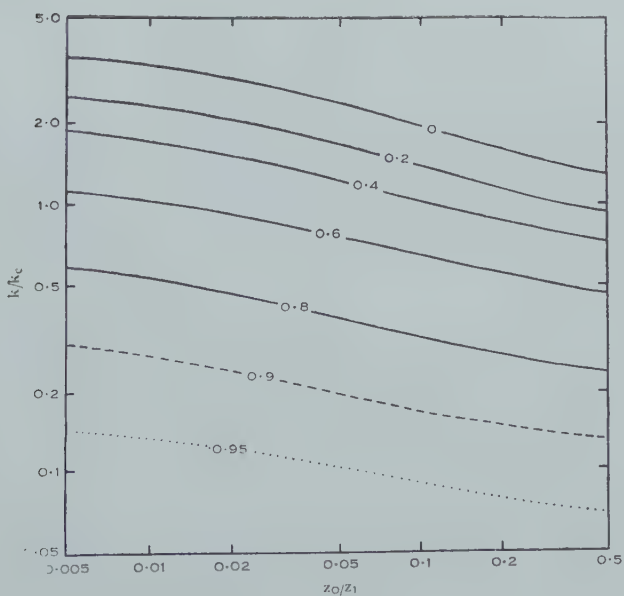


Fig. 5.—Mixing factor for temperature ($T'/T' = n_2^*/n_2$).

the value $1/25$ being taken as it is roughly the median for the integral I_3 . This means that the numerical results (25)–(29) are approximately corrected for mixing by multiplication by the factor

$$\int_0^\infty f\left(\frac{k}{k_c}\right) \frac{n_i\left(\frac{k}{k_c}, \frac{1}{25}\right)}{n_i\left(\frac{1}{25}\right)} d\left(\frac{k}{k_c}\right). \quad \dots\dots\dots (32)$$

It is seen from Figures 4 and 5 that the mixing factor for w always exceeds the mixing factor for T' . It follows that the mixing factors for σ_T^2 , F , and σ_w^2 are in ascending order of magnitude.

As predicted, the mixing factors for F and σ_T show no considerable variation with C and z_1 in the data of Tables 4 and 6, and their mean values of $1/4 \cdot 7$ and $1/3 \cdot 1$ indicate, on reference to (32) and Figures 4 and 5, that the dominant part of the $f(k/k_c)$ distribution is at k/k_c between about 1 and 2. The dominant mixing rate at z_1 is derived accordingly from (11) and (7) as

$$k=1 \text{ to } 2\sqrt{Cz_1^{-4/3}}=1 \text{ to } 2\sqrt{\frac{g}{T}\left|\frac{\partial T}{\partial z}+\Gamma\right|}. \quad \dots\dots (33)$$

This expression provides a yardstick for the time of response of sensing elements and recorders necessary for accurate measurement of heat flux and temperature variations.

VI. EXTENSIONS OF THE THEORY

(a) Allowance for Descending Air

Throughout the development of the theory, consideration has been confined to ascending elements. It has been shown in Section III (c) that for an element, starting from rest, to penetrate a finite distance it must be absolutely buoyant ($k < k_c$) at its level of origin; and it will then continue upwards until, through mixing, it is finally brought to rest at some level above that at which it ceases to be absolutely buoyant.

Descending elements, though subject to the same mixing equations as the ascending elements, encounter the differing environmental conditions in the reverse order, and this profoundly alters their behaviour. An element descending with a temperature deficit, if absolutely buoyant in the negative sense at one level, will be so at all lower levels. Under mixing and buoyancy alone, the velocity and temperature deficit would increase indefinitely. Near the heated boundary of a semi-infinite medium, therefore, the factors limiting the motion of the ascending and descending air are not the same; practical evidence comes from the work of Ramdas and colleagues (Ramdas 1953) who have observed a striking difference in character between the motion of the ascending and the descending air. Whereas the ascending elements are restrained by mixing, the principal constraint on the descending air must be exercised by some other influence which is, almost certainly, the presence of the boundary itself.

In order to describe in detail the behaviour of the descending air it would be necessary to formulate the effects of this influence, but for present purposes this can be avoided provided that the influence may be assumed to be independent of the size of the element on which it operates. The descending air may then, as a reasonable working approximation, be treated as a single mass, of uniform velocity and temperature at any given level. Using suffixes a and d for ascending and descending air, the condition for continuity of mass is then

$$-pw_d = q\bar{w}_a, \quad (p+q=1)$$

where p , q denote the respective mass weighting fractions. Since T' is defined as difference from the average temperature of the *whole* environment, there is also the condition

$$-pT'_d = q\bar{T}'_a.$$

The overall average of wT' is then

$$\begin{aligned} q(\bar{wT'})_a + pw_d T'_d &= q(\bar{wT'})_a + \frac{q^2}{p} \bar{w}_a \bar{T}'_a \\ &= \frac{T_e}{g} C^{3/2} z_1^{\frac{4-3\delta}{2}} \left(qN_3 + \frac{q^2}{p} N_1 N_2 \right), \end{aligned}$$

while the overall average of T'^2 is

$$\begin{aligned} q\bar{T'^2}_a + pT'^2_d &= q\bar{T'^2}_a + \frac{q^2}{p} \left(\bar{T}'_a \right)^2 \\ &= \left(\frac{T_e}{g} C z_1^{1-\delta} \right)^2 \left(qN_3^2 + \frac{q^2}{p} N_2^2 \right), \end{aligned}$$

and similarly for w'^2 .

Whereas the numbers N_i have been shown to depend on δ only, it may appear *a priori* that p and q may depend also on the other non-dimensional combination of the basic variables, that is, on $Cz_1^{1-\delta}/g$. But writing the expression for the heat flux as

$$F \propto z_1^{\frac{4-3\delta}{2}} \left\{ \frac{1}{p} N_1 N_2 + (N_3 - 2N_1 N_2) + p(N_1 N_2 - N_3) \right\}$$

it is seen that terms in p , 1, and $1/p$ occur in F so that no relation of the form

$$p \propto \left(\frac{Cz_1^{1-\delta}}{g} \right)^\alpha$$

can give F independent of z_1 , other than $\alpha=0$. The condition that the heat flux shall be constant with height therefore requires both that $\delta=4/3$ and that p and q shall be independent of z_1 ; the ratio of the ascending and descending masses remains constant through the layer of constant F . The relations (17), (20), and (21) giving F , σ_T , and σ_w in terms of C and z_1 then remain unaffected apart from a modification of the numerical factors.

(b) *The Effect of Superimposed Temperature Fluctuations*

In the development of the theory in Section III (b), all elements have been treated as starting from rest at the local temperature of the environment. The superimposition of temperature fluctuations at the level of origin will in general modify the heat flux and other statistics of the temperature field (Priestley and Swinbank 1947). The effect has been established within the framework of the classical mixing length theory and requires re-examination on the basis of the present theory in which, unlike the former, the influence of elemental buoyancy is taken into account.

Let T'_0 denote the representative value for the temperature anomaly at the starting level. This gives rise to an additional term gT'_0/T_e on the right of (9), and hence to

$$\frac{gT'_0}{T_e}(z_1 - z_0) \dots\dots\dots (34)$$

on the right of (10). To proceed further it is necessary to know the manner in which T'_0 depends on lapse rate C and on starting height z_0 .

Progress can be made if, as seems reasonable on intuitive grounds, it be assumed that T'_0 will be proportional to the temperature fluctuations otherwise present at the level z_0 , that is, from (21)

$$T'_0 \propto \frac{T_e}{g} Cz_0^{1-\delta}, \dots\dots\dots (35)$$

and the additional term (34) on the right of (10) then becomes proportional to

$$Cz_1^{1-\delta}(z_1 - z_0). \dots\dots\dots (36)$$

It may be seen from the modified form of (10) or (13) that the addition of (36) does not affect the previous conclusion that the solution of (10) is of the functional form

$$w = C^{\frac{2-\delta}{2}} z_1^{\frac{2-\delta}{2}} n_1,$$

although n_1 will not have the same numerical values as when the term (36) is omitted. It also follows that the modification to (14) consists simply of the addition of T'_0 to the right-hand side whence, invoking (35) again, (15) and hence (21) remain unaffected in functional form. It is thereby shown that the hypothesis (35) is internally consistent.

On this basis, therefore, the main arguments of the paper extend immediately to embrace the effects of imposed temperature fluctuations, and the temperature profile law and relations derived therefrom would remain unaltered in functional form. If the imposed fluctuations are considerable and obey some form other than (35), the mathematics become intractable and no general solution has been found. Solutions have, however, been found for the special case when T'_0 and the lapse rate (either super- or sub-adiabatic) are constant with height; these present features of novel interest which will be described in a separate paper.

VII. ACKNOWLEDGMENTS

The author is grateful to his colleagues in the Section of Meteorological Physics for helpful discussion of a previous draft of the paper and for allowing presentation in Tables 4 and 6 of some of their experimental results, in advance of full publication; particularly to Mr. E. L. Deacon for his knowledge of temperature profiles and the analysis summarized in Table 1. Valuable assistance was received from Mrs. J. P. Wilson with the computations of Section V (b).

VIII. REFERENCES

- BACHELOR, G. K. (1953).—*Quart. J. R. Met. Soc.* **79**: 224-35.
 BEST, A. C. (1935).—*Geophys. Mem. Lond.* No. 65.
 BOSWORTH, R. C. L. (1952).—"Heat Transfer Phenomena." (Assoc. Gen. Publ.: Sydney.)
 CHANDRA, K. (1938).—*Proc. Roy. Soc. A* **164**: 231-42.
 DEACON, E. L. (1948).—*Quart. J. R. Met. Soc.* **74**: 410-11.
 DEACON, E. L. (1953).—*Geophys. Mem. Lond.* No. 91.
 FISHENDEN, M., and SAUNDERS, O. A. (1932).—"The Calculation of Heat Transmission." (H.M. Stationery Office: London.)
 FLOWER, W. D. (1937).—*Geophys. Mem. Lond.* No. 71.
 GERHARDT, J. R., JEHN, K. H., GUILD, W. R., and STALEY, R. C. (1948).—*Elec. Eng. Res. Lab. Univ. Texas Rep.* No. 22.
 GERHARDT, J. R. (1949).—*Elec. Eng. Res. Lab. Univ. Texas Rep.* No. 29.
 DE GRAAF, J. G. A., and VAN DER HELD, E. F. M. (1953).—*Appl. Sci. Res. Hague A* **3**: 393-409.
 JOHNSON, N. K., and HEYWOOD, G. S. P. (1938).—*Geophys. Mem. Lond.* No. 77.
 PASQUILL, F. (1949).—*Proc. Roy. Soc. A* **198**: 116-40.
 PRIESTLEY, C. H. B. (1953).—*Aust. J. Phys.* **6**: 279-90.
 PRIESTLEY, C. H. B., and SWINBANK, W. C. (1947).—*Proc. Roy. Soc. A* **189**: 543-61.
 RAMDAS, L. A. (1953).—*Proc. Ind. Acad. Sci.* **37**: 304-16.
 RIDER, N. E., and ROBINSON, G. D. (1951).—*Quart. J. R. Met. Soc.* **77**: 375-401.
 ROBINSON, G. D. (1950).—*Cent. Proc. R. Met. Soc.*: 26-9.
 SUTTON, O. G. (1948).—*Quart. J. R. Met. Soc.* **74**: 13-30.
 SUTTON, O. G. (1953).—"Micrometeorology." (McGraw-Hill: New York.)
 SWINBANK, W. C. (1950).—*J. Met.* **8**: 135-45.

VERTICAL HEAT TRANSFER FROM IMPRESSED TEMPERATURE FLUCTUATIONS

By C. H. B. PRIESTLEY*

[Manuscript received November 27, 1953]

Summary

The ability of buoyant elements to carry heat upwards through a stably stratified fluid depends on their rate of mixing and hence on their size. The largest and smallest elements are both relatively ineffective and there exists an optimum intermediate size yielding a maximum value of the buoyant heat flux for a given intensity of temperature disturbance.

For a layer of uniform unstable stratification the heat flux increases progressively with size of element and there is no theoretical upper limit apart from that set by the depth of the unstable layer.

The distribution, with respect to element size, of the intensity of temperature fluctuations impressed by external influences is modified by the effects of buoyancy and mixing, and relations are derived between the modified and unmodified distributions.

I. INTRODUCTION

It is proposed to examine theoretically the vertical flow of heat which results when a layer of fluid is subject to the continual creation of hot elements within it. Whenever turbulence occurs in a thermally stratified liquid, the action of pressure forces which are dissociated from the temperature fluctuations will bring about a state of affairs with elements at rest differing in temperature from their surroundings, but these differences have not normally been allowed for in constructing the equations of heat transfer. The problem therefore has quite general significance, but it becomes of special importance in meteorology where at least two further mechanisms for the creation of "hot spots" may be identified. The natural surface of the Earth is uneven in its physical properties, and so when heated or cooled is subject to local variations in surface temperature which are communicated to the overlying air; the second mechanism occurs when convective clouds, generated in an unstable layer, penetrate a stable or less unstable layer above in which the heat flow requires to be studied.

The buoyant motions resulting from the creation of hot spots will account for a component of heat flux, F_H , whose sense will always be upwards but whose magnitude may depend on the thermal stratification of the fluid. The need is to express F_H in terms of the stratification, of the size of the heated elements, and of the intensity T'_0 of the impressed† temperature fluctuations. This

* Section of Meteorological Physics, C.S.I.R.O., Melbourne.

† The word is used to denote fluctuations arising from causes other than the motion or mixing of the elements themselves.

problem will here be treated for a layer in which the temperature gradient and the intensity T'_0 and frequency of impressed fluctuations (i.e. the number appearing per unit height per unit time) are constant with height and time.

II. BASIS OF THE TREATMENT

A formulation has been given earlier (Priestley and Swinbank 1947), but it is possible to treat in greater detail by invoking the model which may most suitably be described as the *open parcel*. This differs from the *closed parcel*, which has hitherto been used in studies of convection and turbulent transfer, in that as it moves it is subject to continuous mixing with its environment. The equations for the vertical motion w and excess temperature T' of an open parcel moving through an environment at rest at temperature T_e are (Priestley 1953)

$$\dot{w} = \frac{g}{T_e} T' - k_1 w, \quad \dots\dots\dots (1)$$

$$\dot{T}' = -w \left(\frac{\partial T_e}{\partial z} + \Gamma \right) - k_2 T', \quad \dots\dots\dots (2)$$

where g and Γ have their usual significance and k_1 and k_2 are the mixing rates. The latter are constant for an individual element but vary, in constant proportion to each other, from one element to another, taking relatively large values for the small and small for the large elements, and so are used in effect to identify the size of parcel under consideration.

T' may be eliminated from (1) and (2) and, assuming T'/T_e is small, the equation of motion of the individual parcel is derived as

$$\ddot{w} + (k_1 + k_2) \dot{w} + \left[\frac{g}{T_e} \left(\frac{\partial T_e}{\partial z} + \Gamma \right) + k_1 k_2 \right] w = 0, \quad \dots\dots (3)$$

an equation with constant coefficients whose solutions are readily obtained. We shall consider first a population of elements all of a given size (given k_1 and k_2) but starting from rest at different levels with temperature excess T'_0 , and derive expressions for the resulting heat flux F_H and r.m.s. temperature fluctuations σ_T at a fixed level in terms of k_1 , k_2 , and T'_0 . The properties of these expressions will then be discussed, with particular reference to their dependence on k_1 and k_2 . This will amount in essence to a discussion of the manner in which the size-distribution functions for F_H and σ_T are related to each other and to the corresponding function for the impressed temperature fluctuations T'_0 .

III. SOLUTION OF THE PROBLEM

When

$$\frac{g}{T_e} \left(\frac{\partial T_e}{\partial z} + \Gamma \right) + k_1 k_2 < 0, \quad \dots\dots\dots (4)$$

that is, when the lapse rate is sufficiently unstable and the element sufficiently large, the motion is absolutely buoyant (Priestley 1953) and both the w and T'

of the individual element ultimately increase exponentially with time. This case will not be considered in this section, but it will be referred to in Section IV (b).

When

$$\frac{g}{T_e} \left(\frac{\partial T_e}{\partial z} + \Gamma \right) + k_1 k_2 > 0, \quad \dots \dots \dots (5)$$

the motion of the individual elements is always bounded, and there are two principal types of motion to consider. Writing

$$\mu^2 = \left| \frac{g}{T_e} \left(\frac{\partial T_e}{\partial z} + \Gamma \right) - \frac{(k_1 - k_2)^2}{4} \right|, \quad \dots \dots \dots (6)$$

then, when the expression inside the modulus is positive, the solution of (3) which satisfies the initial conditions $w=0$, $T'=T'_0$ is of the oscillatory form

$$w = A e^{-\frac{1}{2}(k_1 + k_2)t} \sin \mu t, \quad \dots \dots \dots (7)$$

whence

$$\frac{T'}{T_e} = \frac{A}{g} e^{-\frac{1}{2}(k_1 + k_2)t} \left(\frac{k_1 - k_2}{2} \sin \mu t + \mu \cos \mu t \right), \quad \dots \dots (8)$$

where $A = (g/\mu)(T'_0/T_e)$. When the expression inside the modulus of (6) is negative, the solution satisfying the same initial conditions is

$$w = A e^{-\frac{1}{2}(k_1 + k_2)t} \sinh \mu t, \quad \dots \dots \dots (9)$$

$$\frac{T'}{T_e} = \frac{A}{g} e^{-\frac{1}{2}(k_1 + k_2)t} \left(\frac{k_1 - k_2}{2} \sinh \mu t + \mu \cosh \mu t \right), \quad \dots (10)$$

with A as before. This is of the asymptotic form.

In deriving the statistical quantities F_H and σ_T^2 at a fixed reference level z_1 , it is recognized that elements of a given size will all be subject to the same motion but will reach z_1 at different stages thereof through having started from different levels z_0 . The contribution to the heat flux is

$$\rho c_p \times \text{average value of } wT',$$

the average being taken over all values of z_0 , weighted according to their probability. The *a priori* probability of z_0 is uniform since conditions, including the frequency of impressed temperature fluctuations, are supposed constant with height, but the recognition that the element has reached the level z_1 affects the *a posteriori* probability of z_0 ; some values are excluded because of the bounded nature of the motion; others, when solutions (7) and (8) apply, must be counted n times where wT' is an n -valued function of $z_1 - z_0$. Since, given the lapse and mixing rates, w and T' depend solely on the difference $z_1 - z_0$, it is mathematically equivalent to keep z_0 fixed and vary z_1 , so that the average wT' may be written

$$\int wT' ds / \int ds,$$

where the integrals are taken over the complete path of an individual element. In this way, and in a similar way for σ_T^2 , is finally derived

$$\frac{F}{\rho c_p} = \int_0^\infty w |w| T' dt \bigg/ \int_0^\infty |w| dt, \quad \dots\dots\dots (11)$$

and

$$\sigma_T^2 = \int_0^\infty |w| T'^2 dt \bigg/ \int_0^\infty |w| dt, \quad \dots\dots\dots (12)$$

where w and T' are as in (7)-(10) and the integral is over the lifetime of the element.

The evaluation of the expressions (11) and (12) is straightforward though laborious, and the following results are obtained. In the case of asymptotic motion (solutions (9) and (10)),

$$\frac{F_H}{\rho c_p} = T_0'^2 \frac{g}{T_e} \cdot \frac{\frac{2k_1}{3}}{\frac{g}{T_e} \left(\frac{\partial T_e}{\partial z} + \Gamma \right) + k_1 k_2 + 2(k_1 + k_2)^2}, \quad \dots\dots\dots (13)$$

$$\sigma_T^2 = \frac{1}{3} T_0'^2 \cdot \frac{\frac{g}{T_e} \left(\frac{\partial T_e}{\partial z} + \Gamma \right) + k_1 k_2 + 2k_1^2}{\frac{g}{T_e} \left(\frac{\partial T_e}{\partial z} + \Gamma \right) + k_1 k_2 + 2(k_1 + k_2)^2}, \quad \dots\dots\dots (14)$$

whence also

$$\frac{F_H}{\rho c_p} = \sigma_T^2 \frac{g}{T_e} \cdot \frac{2k_1}{\frac{g}{T_e} \left(\frac{\partial T_e}{\partial z} + \Gamma \right) + k_1 k_2 + 2k_1^2}. \quad \dots\dots\dots (15)$$

In the case of oscillatory motion (solutions (7) and (8)), the expressions on the right of (13) and (14) must be multiplied by the factor

$$\coth \frac{3(k_1 + k_2)\pi}{4\mu} \bigg/ \coth \frac{(k_1 + k_2)\pi}{4\mu}$$

and the relation (15) continues to hold.

In discussing the results it will be convenient to define a function f_1 in the relation between heat flux F_H and impressed temperature fluctuations by

$$\frac{F_H}{\rho c_p} = \frac{g}{T_e} T_0'^2 f_1,$$

where f_1 contains the dependence on lapse rate and, through k_1 and k_2 , on element size; similarly the relation between heat flux and r.m.s. temperature fluctuations at the same level may be described by the function f_2 defined by

$$\frac{F_H}{\rho c_p} = \frac{g}{T_e} \sigma_T^2 f_2.$$

Of these two functions, the greater practical interest attaches to f_2 , since it is possible to measure both σ_T^2 and the total heat flux, of which F_H is a component, whereas T_0' is not easily measurable. There is, however, some interest in the quantity $\sqrt{f_1/f_2}$ which is equal to σ_T/T_0' and so represents the factor by which the impressed temperature fluctuations are reduced when records of temperature are taken at a fixed level.

The discussion will deal separately with stable and unstable stratifications, so that in writing

$$\lambda^2 = \left| \frac{g}{T_e} \left(\frac{\partial T_e}{\partial z} + \Gamma \right) \right|, \quad \dots\dots\dots (16)$$

there need be no confusion through use of the modulus. It will then be seen that both the criteria for the solutions and the values of λf_1 and λf_2 appertaining thereto depend only on the ratios k_1/k_2 , k_1/λ , and k_2/λ . The results may therefore be presented in completeness by diagrams in which the dimensionless quantities λf_2 and $\sqrt{f_1/f_2}$ are displayed as functions of k_1/k_2 ($=\beta$) and k_2/λ ($=\xi$).

IV. SIGNIFICANCE OF THE RESULTS

(a) *Stably Stratified Medium*

Although either asymptotic or oscillatory motion can occur at stable lapse rates, the form of (15) and hence of f_2 is independent of the mode of motion occurring. With the notation adopted, (15) becomes

$$\lambda f_2 = \frac{2\beta\xi}{1 + \xi^2(2\beta^2 + \beta)}. \quad \dots\dots\dots (17)$$

Isopleths of λf_2 , as a function of β and ξ , are shown in Figure 1.

The most significant feature is marked by the broken line, whose equation is

$$\xi^2 = \frac{1}{2\beta^2 + \beta}, \quad \dots\dots\dots (18)$$

along which λf_2 attains its maximum values. As stated in Section II, the ratio β between the mixing rates for momentum and sensible heat is to be regarded as independent of the size of element, and as determined by factors not here under consideration. Special interest attaches to the case in which these mixing rates are equal, $\beta=1$, when from the above

$$k_1^2 = k_2^2 = \frac{\lambda^2}{3} = \frac{1}{3} \frac{g}{T_e} \left(\frac{\partial T_e}{\partial z} + \Gamma \right) \quad \dots\dots\dots (19)$$

is derived as the condition for maximum heat flux.

The interpretation of this result is that, at a given stable lapse rate, there exists an optimum size of element which is most efficient for the buoyant transport of heat, both very large and very small elements being relatively inefficient in this respect. The smallest elements are inefficient on account of their rapid rate of mixing and consequent short life; k_1 and k_2 are large and the term in $\exp \{ -\frac{1}{2}(k_1 + k_2)t \}$ dominates the solutions (7)-(10), whichever pair is applicable. The reason for the inefficiency of very large elements, to which the oscillatory

solution applies, is that the damping becomes zero so that the motion approaches the simple harmonic with large amplitudes for both w and T' , but the phase difference approaches $\pi/2$ and little heat flux results.

It is of interest to remark that, for solutions of the type under discussion, the vanishing of the mean product wT' while the motion remains finite requires *both* that the phase difference of the periodic part shall be $\pi/2$ *and* that the damping be negligible. For $k_1=k_2$ the first condition is satisfied for elements of all sizes, the second only for the largest ones.

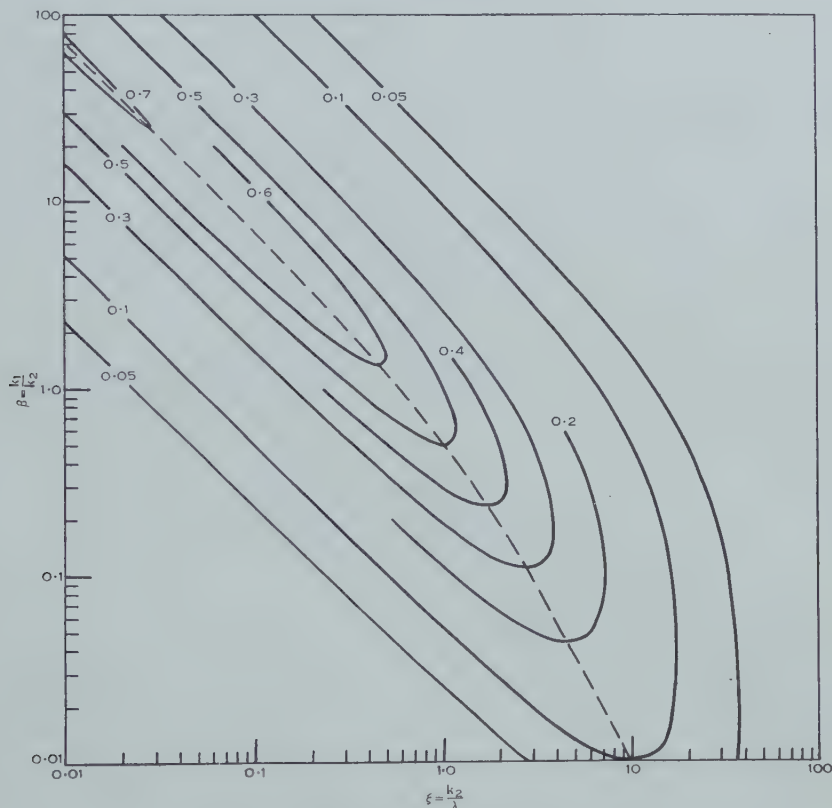


Fig. 1.—Isopleths of λf_2 in stable lapse rates.

For the optimum size of element, or mixing rate, given by (19), (15) takes the value

$$F_H = \frac{\rho c_p \sigma_T^2}{\sqrt{\frac{3T_e}{g} \left(\frac{\partial T_e}{\partial z} + \Gamma \right)}}, \quad \dots \dots \dots (20)$$

which represents therefore an upper limit to the heat flux due to buoyancy, given σ_T and the temperature gradient. It may in practice prove a generous upper limit, since only a small fraction of elements may be close to the optimum size.

Reference to some actual magnitudes observed in the atmosphere is instructive. Fairly strong stable stratification at a height of 1 m above ground may be represented by $\partial T_e/\partial z + \Gamma = 25 \times 10^{-4} \text{ }^\circ\text{C/cm}$, for which the optimum k from (19) is given as $1/20 \text{ sec}^{-1}$. Representative values of σ_T observed under these conditions over a uniform site at Edithvale, Victoria, are $0.2 \text{ }^\circ\text{C}$, whence the upper limit to the heat flux due to buoyancy is about 1 mW/cm^2 . This is of the order of magnitude of the *total downwards* heat flux under such conditions, with winds of $1\text{--}2 \text{ m/sec}$, as measured by Swinbank (1952).

For a layer above convection cloud, into which that cloud might penetrate, we might take $\partial T_e/\partial z + \Gamma = 10^{-5} \text{ }^\circ\text{C/cm}$, whence the optimum k would be $1/5 \text{ min}^{-1}$. With σ_T of about $0.5 \text{ }^\circ\text{C}$ (Byers and Braham 1949), the upper limit of F_H would be about 100 times the above. There is therefore the possibility of a considerable upward flow of heat through statically stable layers in the free atmosphere.

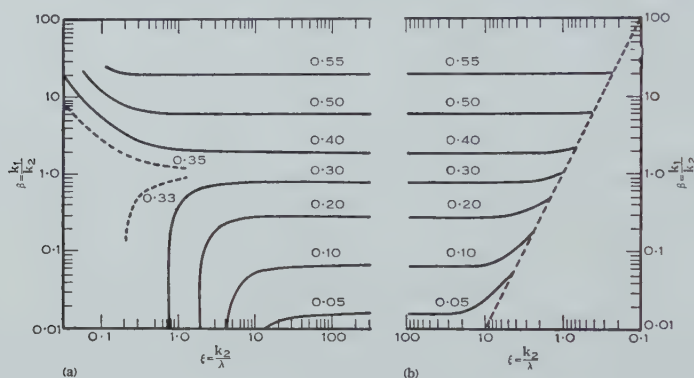


Fig. 2.—Isopleths of the reduction factor for temperature fluctuations σ_T/T_0' . (a) Stable lapse rates, (b) unstable lapse rates.

The values of the reduction factor $\sqrt{f_1/f_2}$ are shown in Figure 2 (a) as a function of β and ξ . When the mixing rates are equal ($\beta=1$) the factor is approximately $\frac{1}{3}$ for elements of all sizes. For the larger elements the factor remains $\frac{1}{3}$ except when the mixing rate for momentum greatly exceeds that for temperature, while for smaller elements the variation of $\sqrt{f_1/f_2}$ with k_1/k_2 is more regular, the two ratios increasing or decreasing in sympathy.

(b) Unstably Stratified Medium

When $(\partial T_e/\partial z + \Gamma)$ is negative, the reduction factor $\sqrt{f_1/f_2}$ takes the same value as before for the smaller elements, and there is little significant change in this value until the limit is reached at which (14) ceases to apply. This is given by $k_1 k_2 = -(g/T_e)(\partial T_e/\partial z + \Gamma)$. The function is shown in Figure 2 (b), the limit being indicated by the broken line.*

* This and the preceding result indicate that, under conditions of practical importance, the simple relation $\sigma_T = \frac{1}{3} T_0'$ will be widely valid.

The function f_2 , which represents the variation of F_H/σ_T^2 , is now

$$f_2 = \frac{1}{\lambda} \cdot \frac{2\beta\xi}{-1 + \xi^2(2\beta^2 + \beta)} \dots\dots\dots (21)$$

As with stable lapse rates the smallest elements (ξ large) can transport little heat, but f_2 increases progressively with size of element until $k_2 = \lambda/\sqrt{\beta}$ at which the element becomes absolutely buoyant, when the relation ceases to apply. f_2 is still finite at this point. The elements will not, however, necessarily be confined to sizes below this critical value. Beyond it, the solutions for w and T' ultimately increase exponentially with time, and the present considerations provide no upper limit to the heat flux which might develop.

The heat flux will not in practice be unlimited since superadiabatic conditions occur only in layers of limited depth, with consequent restriction of a different type on the development of w and T' . In the layer close to a heated surface further restrictions reside in the motion-inhibiting presence of the boundary and consequent curvature of the temperature profile; the problem of heat transfer in this layer is discussed in a paper published concurrently with this (Priestley 1954).

V. REFERENCES

- BYERS, H. R., and BRAHAM, R. R. (1949).—"The Thunderstorm." (U.S. Weather Bureau: Washington, D.C.)
- PRIESTLEY, C. H. B. (1953).—*Aust. J. Phys.* **6**: 279-90.
- PRIESTLEY, C. H. B. (1954).—*Aust. J. Phys.* **7**: 176.
- PRIESTLEY, C. H. B., and SWINBANK, W. C. (1947).—*Proc. Roy. Soc. A* **189**: 543-61.
- SWINBANK, W. C. (1952).—U.S.A.F. Geophys. Res. Paper No. 19, pp. 355-64.

DEVIATIONS FROM MATTHIESSEN'S RULE FOR COLD-DRAWN WIRES

By G. J. OGILVIE* and W. K. CLOTHIER†

[Manuscript received October 7, 1953]

Summary

An A.C. bridge method of high accuracy has been developed to measure deviations from Matthiessen's rule. Deviations have been shown to exist in cold-drawn wires of copper, aluminium, bronze, and 80/20 brass.

I. INTRODUCTION

In a recent investigation (Boas and Nicholas 1953) it was found that the slope $\Delta\rho/\Delta T$ of the resistivity *v.* temperature curves of cold-drawn wires was smaller by a few per cent. than that of annealed wires in the cases of 75/25 brass and an aluminium bronze. However, for six other metals and alloys the difference in slope was within the limits of experimental accuracy. The purpose of the present work was to check the previous results for brass and aluminium bronze and to find out whether pure metals observe Matthiessen's rule strictly, or also show a change in $\Delta\rho/\Delta T$ on cold-drawing.

Boas and Nicholas measured the resistances of their wires using a Kelvin double bridge, the wires being immersed one at a time and individually measured in melting ice and liquid oxygen. In the present experiments an A.C. bridge method was used which offers the advantage of high sensitivity and elimination of errors due to thermo-e.m.f.'s. The high sensitivity available enabled the use of small measuring currents, thus avoiding errors due to temperature rise in the wires being measured. Moreover, both wires were immersed together in the same bath ensuring that their temperature was almost the same and were measured almost simultaneously so that errors in the measurement of the temperature of the bath are not critical.

II. EXPERIMENTAL

(a) *Preparation of Specimens*

The materials used (conductivity copper, 80/20 brass, and an aluminium bronze) and their compositions were the same as those used in the previous experiments. They were originally in the form of factory annealed wires of approximately 4 mm diameter and were given the following heat treatment prior to drawing. The copper wires were annealed *in vacuo* at 600 °C for 1 hr. The 80/20 brass wire was sealed in coal gas at atmospheric pressure, the aluminium bronze specimens were sealed in coal gas at approximately 1 mm Hg; they

* Division of Tribophysics, C.S.I.R.O., University of Melbourne.

† Division of Electrotechnology, C.S.I.R.O., University Grounds, Sydney.

were annealed at 600 °C for 1 hr. The wires were then drawn on a power draw bench at room temperature. The drawing speed was maintained as constant as possible at about 4½ ft/min. A specimen was taken from each wire, one-half of which was left in the cold-drawn condition and placed in a refrigerator, while the other half was annealed under the same conditions as those given for the primary material.

It was found that the diameter of a wire could increase from beginning to end by more than 1 per cent. if only short lengths (2 or 3 ft) were drawn. Since most of the change occurred in the first 8 ft of the wire no specimens were taken from the first 12 ft of wire to pass through the die. The cause of these diameter variations is not completely understood. It seems likely that heating of the die and the consequent changes in the properties of the lubricant used could account for most of the diameter variation. Along the specimen length (approximately 60 cm) the diameter variation was less than 3×10^{-5} in.

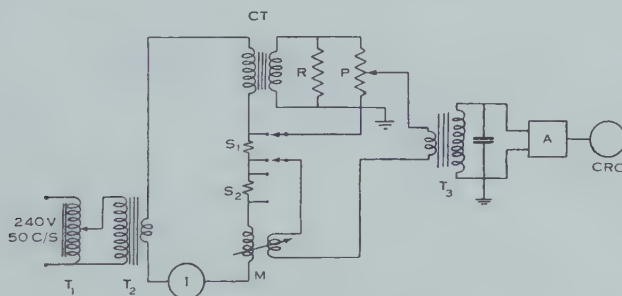


Fig. 1.—A.C. bridge circuit for resistance measurements.

(b) Apparatus

The resistance measurements were made by an A.C. bridge method similar to that used by Broom and Clothier (1952). The circuit is shown in Figure 1. Current at 50 c/s is supplied from the variable auto-transformer T_1 and isolating transformer T_2 , to the specimens S_1 and S_2 and also to the primaries of the current transformer CT and variable mutual inductor M . The secondary current of CT flows through a low value four-terminal resistor R , the potential terminals of which are connected to a Thomson-Varley potential divider P . The settings of P and M are adjusted until the detector indicates balance. The balance detector comprises a tuned step-up transformer T_3 , tuned amplifier A and cathode-ray oscillograph CRO .

The specimen resistance is given in terms of the ratio of CT , the value of R , and the setting of P at balance. The value of M is very small and it does not enter significantly into the determination of R . Relative values of specimen resistance are given directly by the relative settings of P . The measuring current in all cases was 0.4 A. Several different values for R were used in the course of the measurements, the value selected being one which ensured adequate reading accuracy on P for the particular specimen pair under test.

Since the resistance of a wire is altered when it is stressed, the comparison of resistivities to the required accuracy is only possible if the wire is not heavily loaded. Moreover, owing to the dimensional changes that occur in the specimen and its support between the upper and lower extremes of temperature used in the test, special precautions must be taken to avoid movement of the contacts relative to the specimen. Therefore, the usual type of specimen holder, in which the wire is held straight in tension and then two knife-edges brought into contact with it, is not satisfactory.

The specimen holder used consisted of a number of "Perspex" plates held parallel by a brass channel. Each plate was drilled with holes, the centres of which were collinear. These holes constrained the specimen wires to lie in a straight line. The knife-edges were supported by the specimen wires themselves; they were formed from U-shaped loops of 20 gauge piano wire, one end being set in a small "Perspex" block which was grooved to hold the specimen wire. The arrangement is shown in Figure 2. A small manipulator was used to place

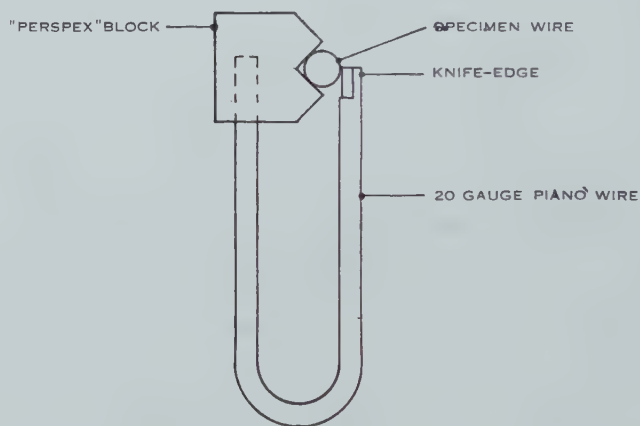


Fig. 2.—Knife-edge assembly.

the knife-edges on the wire so that the least possible bending and surface damage was caused. The separation between the knife-edges, which were about 15 cm apart, was measured with a cathetometer which could be read to 0.001 cm. The wires were connected in series using a bridging piece at one end. The specimen assembly was sufficiently small to fit comfortably into a medium size parallel-sided Dewar flask which served as a constant temperature bath for the measurements at low temperature. The measurements were carried out at room temperature and at liquid oxygen temperature.

The temperature at which the measurements were made was not of prime importance, but it was necessary that the temperatures of the specimen wires be equal to high precision (better than 0.01 °C). To achieve this in the water-bath at room temperature efficient stirring was necessary. Preliminary measurements established that the conductivity of tap water was not high enough to interfere with the electrical measurements. As a precaution, however, distilled water was used. The experiments were performed in a constant

temperature room and, since the bath was large and at the same temperature as the surrounding air, temperature drifts were extremely slow.

The low temperature bath was liquid oxygen. It was necessary to leave the specimen assembly in the bath for at least 10 min before measurements were started to ensure temperature equilibrium. In addition, all metal parts of the specimen assembly had to be completely covered by the liquid oxygen, otherwise irregular temperature changes caused variations in the resistance of the wires. Under these conditions the temperature difference between the wires was very small. This was shown by stirring the bath and finding that, with two identical wires in the holder, the ratio of the resistances remained constant. The observed change in resistance of each wire would have been accounted for by a fall in temperature of approximately 0.1°C .

As a check, the resistances were frequently remeasured at room temperature after the specimens had been in the liquid air bath, and then again in the low temperature bath. After correcting for temperature changes in the room temperature bath it was found that the resistance of the wires, after immersion in liquid oxygen, changed by less than 0.02 per cent.

On removing the specimen assembly from the room temperature bath, prior to immersion into liquid oxygen, it was passed through two successive alcohol baths and then dried in an air stream. It was necessary to ensure that no water was left under the knife edges, since the presence of ice can cause trouble by lifting the knife-edges out of contact with the wire. For the same reason it was found necessary to avoid icing of the contacts in transferring the specimen assembly from the liquid oxygen bath to the water-bath. Failure to observe this precaution caused large changes in the measured resistance of the wires, due, it is believed, to displacement of the contacts. In order to avoid temperature changes in the room temperature bath, the specimen assembly was brought nearly to the correct temperature in an intermediate water-bath before placing it in the room temperature bath for measurement.

After the completion of the electrical tests the portion of the specimen between the potential points was cut out and its ends were squared. It was weighed and its length was measured whilst held straight in a glass capillary tube.

III. RESULTS

The experimental results are summarized in Table 1, which also gives the results of Boas and Nicholas for comparison. The deviations from Matthiessen's rule found in the present work lie within the rather large limits of error of the previous experiments. Further, such deviations have now been shown to exist in copper. The specimens investigated all showed "negative" deviations from Matthiessen's rule, that is, $(\Delta\rho/\Delta T)_D$ is less than $(\Delta\rho/\Delta T)_A$. Absolute values of the resistivity can be calculated only if the cross sections of the wires are known. In uniform wires of small diameter the most accurate method of determining cross section is to measure the mass and length of the wires. This method requires a knowledge of the density. As this information is not available to the necessary precision, the products of the resistivity and the density are

tabulated. In the determination of relative resistivity it is assumed that the densities of the deformed and annealed wires forming a specimen pair are the same.

TABLE I
SUMMARY OF RESULTS

1	2	3	4	5*	6*	7	8
Metal	State	Dia- meter (in.)	Logar- ithmic Strain	$(\rho \times d) \times 10^{-5}$ at Room Temp.	$(\rho \times d) \times 10^{-5}$ at Liquid Oxygen Temp.	$\left[\frac{(\Delta\rho/\Delta T)_D}{(\Delta\rho/\Delta T)_A} - 1 \right] \times 100$	
						Present Work	Boas and Nicholas
80/20 Brass	Deformed	0.0359	3.19	5.632	4.101	-0.02 ± 0.04	
	Annealed	"	"	4.721	3.190		
	Deformed	"	"	5.631	4.100	-0.06 ± 0.04	-0.06 ± 0.66
	Annealed	"	"	4.721	3.189		
Alu- minium bronze	Deformed	0.0159	4.97	11.275	9.806	-2.2 ± 0.04	
	Annealed	"	"	8.639	7.137		
	Deformed	0.0285	3.81	11.238	9.766	-1.7 ± 0.04	-1.98 ± 1.69
	Annealed	"	"	8.654	7.156		
	Deformed	"	"	11.238	9.755	-1.9 ± 0.04	
	Annealed	"	"	8.654	7.142		
Copper	Deformed	0.0403	3.11	11.153	9.677	-1.7 ± 0.04	
	Annealed	"	"	8.607	7.104		
	Deformed	0.0179	3.93	1.5559	0.3096	-0.29 ± 0.04	
	Annealed	"	"	1.5107	0.2608		
	Deformed	0.0201	3.70	1.5543	0.3076	-0.29 ± 0.04	
	Annealed	"	"	1.5112	0.2607		
	Deformed	0.0253	3.24	1.5472	0.3052	-0.19 ± 0.04	0.02 ± 0.30
	Annealed	"	"	1.5046	0.2602		
	Deformed	0.032	2.77	1.5428	0.3033	-0.22 ± 0.04	
	Annealed	"	"	1.5019	0.2595		
	Deformed	0.0403	2.31	1.5436	0.3009	-0.12 ± 0.04	
	Annealed	"	"	1.5038	0.2597		

* The values for the product of the resistivity and the density, $\rho \times d$, given in columns 5 and 6 have been rounded off after calculating column 7.

In the table ρ is the mass resistivity. To convert to volume resistivity thermal expansion must be allowed for. Precise figures are not available, but it is clear that

$$\left[\frac{(\Delta\rho/\Delta T)_D}{(\Delta\rho/\Delta T)_A} - 1 \right] \times 100$$

is not appreciably altered in the cases of copper and aluminium bronze. For 80/20 brass, however, the deviation will be opposite in sign.

IV. ERRORS

- The known sources of error are as follows.

(i) *Uncertainty in the Length of Specimen between the Potential Knife-edges.*—It was required that the wire be unstrained when resistance measurements were being made. On the other hand, to determine the specimen length accurately with a cathetometer, it was necessary that the wire be straight. In the specimen holder, the wire was held very nearly straight. However, for several of the wires the residual irregularities were sufficient to cause errors of about 2 in 10^4 .

(ii) *Uncertainties in Temperature.*—Preliminary measurements established that there could be a temperature difference of about 0.1°C between the liquid oxygen bath and the specimen wires, whilst the difference in temperature between the wires was considerably less than 0.01°C . Errors due to temperature measurement or to temperature inequality were, therefore, negligible.

(iii) *Errors in Resistance Measurements.*—The absolute accuracy of resistance measurement was better than 2 in 10^4 ; the accuracy of intercomparison of any specimen pair was about 3 in 10^5 and the sensitivity of measurement was such that a variation of 1 in 10^5 could easily be detected.

(iv) *Inhomogeneity of Specimens.*—Using a short distance between the potential contacts, resistance measurements were made at various positions along a few of the specimen wires. Variations of unknown origin were found. Although these variations would be important in determining absolute resistivities, their effect in this determination of the deviations from Matthiessen's rule is very small.

(v) The length of the wires was measured to an accuracy of $\pm 0.005\text{ mm}$ in a length of about 15 cm. Their weights were determined to $\pm 0.00002\text{ g}$; the minimum weight being greater than 0.2 g.

The overall accuracy is sufficient to give

$$\left[\frac{(\Delta\rho/\Delta T)_D}{(\Delta\rho/\Delta T)_A} - 1 \right] \times 100,$$

the deviation from Matthiessen's rule, to an accuracy of ± 0.04 .

V. DISCUSSION

The higher accuracy of the present results is due to the improved technique of measurement and to the recognition of the serious variation in diameter occurring in the first few feet of a drawn wire. This fact sets an important limit on the reliability of all previous results and throws doubt on the conclusions drawn from them (Rutter and Reekie 1950).

Two conclusions can be drawn from the present results. Firstly, deviations from Matthiessen's rule have been found with the materials investigated. The lower accuracy of their method and not the absence of the effect prevented Boas and Nicholas finding significant deviations in copper. Secondly, all the deviations are of the same sign, that is,

$$\left[\frac{(\Delta\rho/\Delta T)_D}{(\Delta\rho/\Delta T)_A} - 1 \right] \times 100$$

is negative in all cases, and this difference increases with the extent of deformation. This is in agreement with the findings of Boas and Nicholas.

In a recent paper Sondheimer (1950) has pointed out that deviations from Matthiessen's rule can be expected at intermediate temperatures. As his calculation holds for monovalent metals only, no comparison of the measured values with his calculation has been made.

Broom (1952) in some experiments on the effect of drawing temperature on the electrical resistance of drawn wires assumed the validity of Matthiessen's rule in the calculation of values of $\Delta\rho/\rho$. The results of the present investigation indicate that this assumption was justified to the accuracy given in his results.

VI. REFERENCES

- BOAS, W., and NICHOLAS, J. F. (1953).—Validity of Matthiessen's rule for cold-worked wires. *Aust. J. Phys.* **6**: 116-21.
- BROOM, T. (1952).—The effect of temperature of deformation on the electrical resistivity of cold-worked metals and alloys. *Proc. Phys. Soc. Lond.* **B 65**: 871-81.
- BROOM, T., and CLOTHIER, W. K. (1952).—The anisotropy of electrical resistivity of cold-drawn wires of some cubic metals and alloys. *Aust. J. Sci. Res.* **A 5**: 119-27.
- RUTTER, J. W., and REEKIE, J. (1950).—The effect of cold-working on the electrical resistivity of copper and aluminium. *Phys. Rev.* **78**: 70-1.
- SONDHEIMER, E. H. (1950).—The theory of the transport phenomena in metals. *Proc. Roy. Soc. A* **203**: 75-98.

THE MULTIPLE SCATTERING OF ELECTRONS AND POSITRONS

By C. B. O. MOHR* and L. J. TASSIE*

[Manuscript received February 9, 1954]

Summary

The angular distribution of the single scattering of 33, 121, and 1065 keV electrons at small angles in gold is calculated and compared with the distributions given by the Born approximation and by the WKB method as used by Molière. The single scattering distribution for 1065 keV electrons is integrated numerically to give mean square angles of multiple scattering, and these are compared with the values given by the various multiple scattering theories. The results are discussed in conjunction with the experimental data for gold and other elements. The discrepancy between theory and the recent experiments with beryllium is shown not to be explained by the use of the Hartree instead of the Thomas-Fermi field. The difference of the root mean square angle for electrons and for positrons is estimated for gold and argon, and its value for argon—the only element for which this difference has been measured—is much less than the observed value.

I. INTRODUCTION

The angular distribution of the scattering of fast electrons and positrons at large angles is now fairly well understood (Bartlett and Watson 1940, Massey 1942, McKinley and Feshbach 1948). The calculations just referred to were carried out for an unscreened Coulomb field, and this is justified, since for large angles of scattering the effect of screening is small for electrons with energy as low as 33 keV for even a heavy element like gold (Mohr and Tassie 1954).

For small angles of scattering, however, screening of the nucleus by the atomic electrons is important. For heavy elements and relativistic energies the Born approximation is not likely to give very accurate results. The WKB method of determining phase shifts has been used together with a method of summing the series of partial waves which should give fairly good results (Molière 1947), but certain approximations have been made and it is of interest to assess the accuracy of the distributions obtained. This may be done by carrying out a detailed phase shift analysis, but particular care has to be taken with the summation of the series of partial waves, for it is extremely slowly convergent at small angles.

Experiments on small angle scattering at high energies almost inevitably involve multiple scattering, any theory of which must be based to some extent on the form of the single scattering distribution. While the multiple scattering distribution does not depend critically on the form of the single scattering distribution, it involves a "screening angle" whose value depends on the form of the atomic field. Some of the multiple scattering theories employ an

* Physics Department, University of Melbourne.

exponentially screened field, others a Thomas-Fermi field, and neither field will lie as close to the true field as the Hartree field. The exact single scattering distribution for gold is therefore used to compute mean square angles, and these are compared with the values given by multiple scattering theories. For this reason only those multiple scattering theories are discussed which give as first approximation a Gaussian distribution whose mean square angle is given by simple analytic formulae.

Finally the difference in the intensity of single scattering of electrons and positrons due to spin—though small at small angles—could give rise to a detectable difference in the width of the respective multiple scattering distributions. There is experimental evidence for such a difference, and its magnitude is therefore investigated theoretically.

II. ANALYTIC THEORIES OF MULTIPLE SCATTERING

The differential cross section for single scattering may be written in the form

$$I(\theta) = (Z^2 \epsilon^4 \operatorname{cosec}^4 \frac{1}{2} \theta / 4 \gamma^2 m^2 v^4) R, \quad \dots \dots \dots (1)$$

where Z is the atomic number of the scattering atom, v is the velocity of the electrons,

$$\gamma = (1 - \beta^2)^{-\frac{1}{2}} \quad \text{with } \beta = v/c,$$

and R is the ratio of the scattering to the relativistic Rutherford scattering.

The chance of occurrence of a single deviation through an angle between θ and $\theta + d\theta$ is given by

$$\begin{aligned} P(\theta) &= 2\pi N t I(\theta) \sin \theta d\theta \quad \dots \dots \dots (2) \\ &= Q R d\theta / \theta^3 \quad \text{for small } \theta, \end{aligned}$$

where

$$Q = 8\pi N t Z^2 \epsilon^4 / \gamma^2 m^2 v^4, \quad \dots \dots \dots (3)$$

N is the number of atoms per c.c. of scattering foil, and t is the foil thickness.

For sufficiently large t there will be a large number of collisions, resulting in an approximately Gaussian distribution with mean square angle given by

$$\theta_{\text{r.m.s.}}^2 \equiv \overline{\theta^2} = \int_0^{\theta_{\text{max.}}} \theta^2 P(\theta) d\theta. \quad \dots \dots \dots (4)$$

It is well known that this integral does not converge to a limit as $\theta_{\text{max.}}$ is increased, but $\theta_{\text{max.}}$ must be appreciably smaller than the root mean square angle if the assumption of a Gaussian distribution is to be justified. An arbitrary "cut-off" to the integral is therefore adopted.

In Williams's (1938) theory of multiple scattering $\theta_{\text{max.}}$ is so chosen that the chance of a deviation through an angle greater than $\theta_{\text{max.}}$ is unity, and therefore we have

$$\theta_{\text{max.}}^2 = \frac{1}{2} Q. \quad \dots \dots \dots (5)$$

It is also assumed that Coulomb scattering ($R=1$) holds down to an angle $\theta_{\min.}$, below which screening is supposed suddenly to reduce the scattering to zero, so that (4) gives

$$\text{Williams :} \quad \theta_{r.m.s.}^2 = Q \ln (\theta_{\max.}/\theta_{\min.}). \quad \dots\dots\dots (6)$$

$\theta_{\min.}$ is chosen to make the mean square angle for Coulomb scattering with a lower cut-off $\theta_{\min.}$ the same as that for scattering by the Thomas-Fermi field, that is, one takes

$$\int_{\theta_{\min.}}^{\theta_{\max.}} \theta^2 (1/\theta^3) d\theta = \int_0^{\theta_{\max.}} \theta^2 (R/\theta^3) d\theta,$$

where the value taken for R is that given by calculating the single scattering by the Thomas-Fermi field using Born's approximation. This finally gives

$$\theta_{\min.} = 1/(65 \cdot 3 \beta \gamma Z^{-1/3}). \quad \dots\dots\dots (7)$$

The somewhat more accurate theory of Goudsmit and Saunderson (1940*a*, 1940*b*) gives, for scattering by the Thomas-Fermi field, an approximately Gaussian distribution with a width given by (Mott and Massey 1949)

$$\text{Goudsmit and Saunderson :} \quad \theta_{r.m.s.}^2 = Q \ln (0 \cdot 64 \theta_{\max.}/\theta_{\min.}). \quad \dots\dots (8)$$

It may be noted that in both theories $\theta_{r.m.s.}$ varies as $t^{1/2} \ln t + \text{const.}$ for a given Z .

The theory of Molière (1948) is based on a value of the screening angle $\theta_{\min.}$ obtained through a calculation of the single scattering by the Thomas-Fermi field using the more accurate WKB method, the value being given by

$$\theta_{\min.} = (1 \cdot 13 + 3 \cdot 76 Z^2/137^2 \beta^2)^{1/2} \theta_0, \quad \dots\dots\dots (9)$$

where

$$\begin{aligned} \theta_0 &= 1/ak = 1/(0 \cdot 885 a_0 Z^{-1/3} k) \\ &= 1/(121 \beta \gamma Z^{-1/3}), \end{aligned}$$

a being the Thomas-Fermi radius of the atom. The theory then gives

$$\text{Molière :} \quad \theta_{r.m.s.}^2 = QB/4, \quad \dots\dots\dots (10)$$

where

$$\begin{aligned} B - \ln B &= b \\ &= 2 \ln (\theta_{\max.}/1 \cdot 08 \theta_{\min.}), \end{aligned}$$

with $\theta_{\max.}$ as previously defined in (5). The appearance of $B/4$ instead of $b/2$ in (10) implies taking a value for the upper limit of the integral (4) which differs from $\theta_{\max.}$ and corresponds to the width of the multiple scattering peak (Bethe 1953).

The theories all contain an addition to the Gaussian distribution which gives the transition to the "single scattering tail" at large angles, and this will slightly alter the width of the multiple scattering peak. We shall for simplicity compare the values of $\theta_{r.m.s.}$ given by the above formulae and those obtained directly by numerical integration in (4), using single scattering distributions calculated in different ways. Molière's theory gives a multiple scattering

distribution narrower than the Gaussian, which occurs as the first term in his distribution; and the total distribution may be fitted—up to an angle where the intensity is $1/e$ of the maximum—by a slightly narrower Gaussian corresponding to a slightly reduced value of B (Hanson *et al.* 1951).

III. CALCULATION OF THE EXACT SINGLE SCATTERING DISTRIBUTION FOR GOLD

The differential cross section $I(\theta)$ is given by

$$I(\theta) = |f|^2 + |g|^2,$$

with

$$2ikf(\theta) = \sum \{ (l+1)(e^{2i\eta_l} - 1) + l(e^{2i\eta_{l-1}} - 1) \} P_l(\cos \theta), \quad \dots (11)$$

and $g(\theta)$ —essentially a spin term—negligible at small angles compared with $f(\theta)$; where $k = 2\pi\gamma mv/h$, and the η are phase shifts of the various order waves which have been calculated accurately for gold (Mohr and Tassie 1954).

The series (11) is very slowly convergent at small angles, and therefore two methods were used for summing it in order to check the accuracy of the results.

Method (a).—The series was summed to the first 30 or 40 terms, and the summation over higher values of l was replaced by an integral, the higher P_l being calculated with high accuracy from the relation

$$P_l(\cos \theta) \simeq (\theta/\sin \theta)^{\frac{1}{2}} J_0((l + \frac{1}{2})\theta). \quad \dots (12)$$

The higher order phases are given with sufficient accuracy by the formula

$$\eta_l \simeq \eta_{l-1} \simeq (\gamma Z_1/ka_0) K_0(\lambda_1(l + \frac{1}{2})/k) + (\gamma Z_2/ka_0) K_0(\lambda_2(l + \frac{1}{2})/k), \quad \dots (13)$$

where the Hartree field of the atom is fitted as closely as possible by the expression

$$(Z_1 e^{-\lambda_1 r} + Z_2 e^{-\lambda_2 r}) \epsilon^2/r \quad \text{with } Z_1 + Z_2 = Z. \quad \dots (14)$$

For gold the values adopted were $Z_1 = 20$, $Z_2 = 59$, $\lambda_1 = 1.3/a_0$, $\lambda_2 = 6/a_0$. It was found necessary to take into account values of l up to 500.

The integrand being oscillatory, the values of the integral between successive zeros of the integrand were obtained by graphical integration. These values constitute a slowly convergent series with alternating signs, the last few terms of which were treated by the Euler transformation (Rosser 1951) in order to speed the convergence.

Method (b).—As a check on the accuracy of the previous method, the following procedure was used in some cases, though it was found in general to be less accurate.

For large l we may take for K_0 in (13) the first term in the asymptotic formula, namely,

$$K_0(x) \simeq (\pi/2x)^{\frac{1}{2}} e^{-x}.$$

Let us denote the resulting expression for η_l by $\eta_{l1} + \eta_{l2}$. One may fit $\sin 2\eta_l$ by the expression

$$2\eta_{l1}(1 - \alpha_1 l^{-\frac{1}{2}}) + 2\eta_{l2}(1 - \alpha_2 l^{-\frac{1}{2}})$$

and $1 - \cos 2\eta_l$ by

$$2\eta_{l1}^2(1 - \beta_1 l^{-\frac{1}{2}}) + 2\eta_{l2}^2(1 - \beta_2 l^{-\frac{1}{2}}) + 4\eta_{l1}\eta_{l2}(1 - \beta_3 l^{-\frac{1}{2}}),$$

where $\alpha_1, \alpha_2, \beta_1, \beta_2$, and β_3 are arbitrarily adjusted constants, the fit being good down to relatively small values of l . The discrepancy at the lower values of l is allowed for in a separate numerical calculation. Substituting these expressions for $\sin 2\eta_l$ and $1 - \cos 2\eta_l$ in (11), replacing the series by an integral, and using

$$(12) \text{ with } l + \frac{1}{2} \text{ replaced by } l, \text{ we require to evaluate the integral } \int_0^\infty l^p e^{-ql} J_0(l\theta) dl$$

for $p = \frac{1}{2}, 0$, and $-\frac{1}{2}$. The value of this integral is given by a hypergeometric series (Watson 1948) which is readily evaluated numerically.

The results of this calculation are shown by the heavy curves marked E_1 and E_2 in Figure 1. Curve E_2 was obtained using the two-term field (14), and E_1 using a one-term field ($Z_1 = 79, Z_2 = 0, \lambda_1 = 3/a_0$) which does not fit the Hartree field for gold so well but which was used as a check on the sensitivity of the results to the precise form of the field. Curve E_H , for 1065 keV electrons, was obtained by using method (a) with the phases calculated accurately for the Hartree field of mercury by Gunnensen (1952).

The undulations in the curves occur near those angles at which either the real or the imaginary part of $f(\theta)$ passes through a zero, and appear to have a real existence. Recent calculations by Hoerni and Ibers (1953) for 40 keV electrons in uranium are found to give a curve of R v. θ which has undulations very similar to those in curves E_1 and E_2 for 33 keV electrons.

IV. COMPARISON OF SINGLE SCATTERING DISTRIBUTIONS FOR GOLD

For comparison, Figure 1 shows curves obtained using the Born approximation, according to which one has

$$f(\theta) = (8\pi^2 m / \hbar^2) \int_0^\infty V(\sin Kr / Kr) r^2 dr, \dots\dots\dots (15)$$

where

$$K = 2k \sin \frac{1}{2}\theta.$$

We insert the form of the potential energy V for the one-term field

$$V = (\gamma Z \epsilon^2 e^{-\lambda r} / r) + (Z^2 \epsilon^4 e^{-2\lambda r} / 2mc^2 r^2),$$

where the second term is a relativistic spin term which contributes appreciably to the scattering only at large angles and has little effect at the angles considered here. Spin terms which have an even smaller effect have been omitted from V . An elementary integration then leads to the result

$$R = \{ (K^2 / \lambda^2 + K^2) + \beta (Z/137) \operatorname{artan} (K/2\lambda) \sin \frac{1}{2}\theta \}^2. \dots (16)$$

The required modification to the formula for the two-term field is obvious.

The curves so obtained for the one- and two-term fields are labelled B_1 and B_2 respectively in Figure 1. As the second term in brackets in (16) is

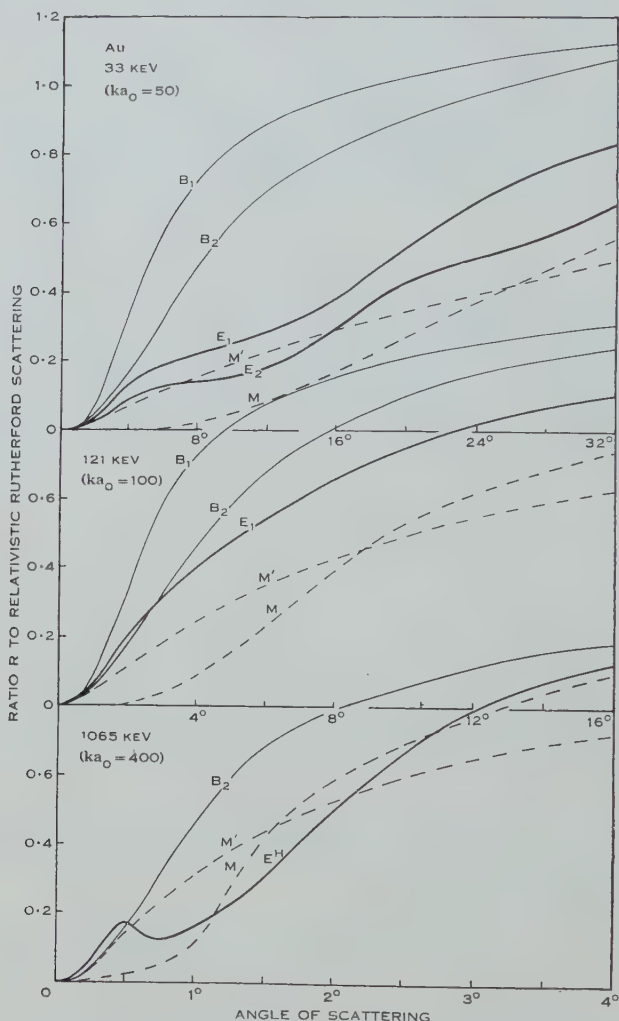


Fig. 1.—Angular distribution of single scattering of electrons in gold. Curves B_1 and B_2 were calculated using the Born approximation with the one-term and two-term potential fields respectively, together with the spin correction term in Z^2 . Curve M was calculated from the formulae (9.1) and (9.3) given by Molière for application to multiple scattering, and M' obtained using Molière's more accurate formula (8.6). Curves E_1 , E_2 , and E_H were calculated using the exact theory in the present paper with the one-term and two-term fields and the Hartree field of mercury respectively.

almost negligible compared with the first term at the angles with which we are concerned, the value of R is practically a function of $k \sin \frac{1}{2} \theta$ only, and this suggests that the sets of curves obtained on the different approximations be

plotted on a scale of θ which is inversely proportional to the value of k . When this is done the Born curves at the different energies are almost identical.

Molière obtains a single scattering distribution, using the WKB method together with a three-term representation of the Thomas-Fermi field, for the limiting cases of $Z/137\beta \rightarrow 0$ (Born approximation) and $Z/137\beta \rightarrow \infty$ (classical approximation). Then, with the aid of an asymptotic formula he produces his empirical interpolation formula (8.6), which gives R for all values of $Z/137\beta$. This leads to the curves labelled M' in Figure 1. For application to the theory of multiple scattering he characterizes the single scattering distribution by a single parameter, the "screening angle" $\theta_{\min.}$ —given by his equation (9.3) (our equation (9)), which implies a single scattering distribution given, somewhat less accurately than by his formula (8.6), by his formula (9.1), namely,

$$R = \theta^4 / (\theta_{\min.}^2 + \theta^2)^2. \quad \dots\dots\dots (17)$$

From this formula are derived the curves marked M . This formula, while more accurate than the first term of (15), has the same form and significance, since $K = k\theta$ for small angles, and $\lambda \sim k\theta_{\min.}$ by a simple application of the uncertainty principle. $\theta_{\min.}$ is the minimum detectable deflection for a field of range $1/\lambda$, and hence for a particle of angular momentum $kh/2\pi\lambda$.

The Born approximation (15) is based on the assumption that the phase shifts are all small compared with unity, whereas in Molière's application of the WKB method this assumption is not made. For a heavy element like gold a large number of the phases are of the order of unity. One therefore finds that the Molière curves M and M' lie closer to the "exact" curves E than do the Born curves B . The supposedly more accurate Molière curve M' , however, falls further and further below the true values of R as the angle increases. This is not surprising, since Molière's value of R approaches 1 always from below, whereas the true value of R becomes greater than 1 for electrons as θ increases (Bartlett and Watson 1940). Furthermore the Molière theory gives the same angular distribution for positrons as for electrons, whereas the value of R for positrons is, at the larger angles where screening is less important, less than that for electrons (Massey 1942). The curve M is a better fit than the curve M' except at the very lowest angles, but it is given by a less accurate formula, so that the closeness of fit is to some degree fortuitous.

As the energy increases the various approximations should improve in accuracy, and one sees that the various theoretical curves for a given energy do approach each other more and more closely with increasing energy.

Finally, the degree of sensitivity of the scattering to the form of the atomic field is shown by the difference between the curves E_1 and E_2 , and between B_1 and B_2 .

V. MEAN SQUARE ANGLE OF MULTIPLE SCATTERING IN GOLD

The values of $\theta_{r.m.s.}$ for 1065 keV electrons predicted by the multiple scattering theories discussed in Section II (formulae (6), (8), (10)) are compared in Figure 2 (curves W , G & S , M_1) with the values obtained by direct numerical integration of the exact single scattering distribution, using (4) (curve E). Also

included in the figure are values obtained by numerical integration of the single scattering distributions given by the Born approximation with the two-term field (curve B), and by the Molière theory (curve M_2). In the latter case the curve M in Figure 1 (given by (17)) is used, as it is this—and not the curve M' —which is given by the use of the screening angle $\theta_{\min.}$ involved in the multiple scattering theory.

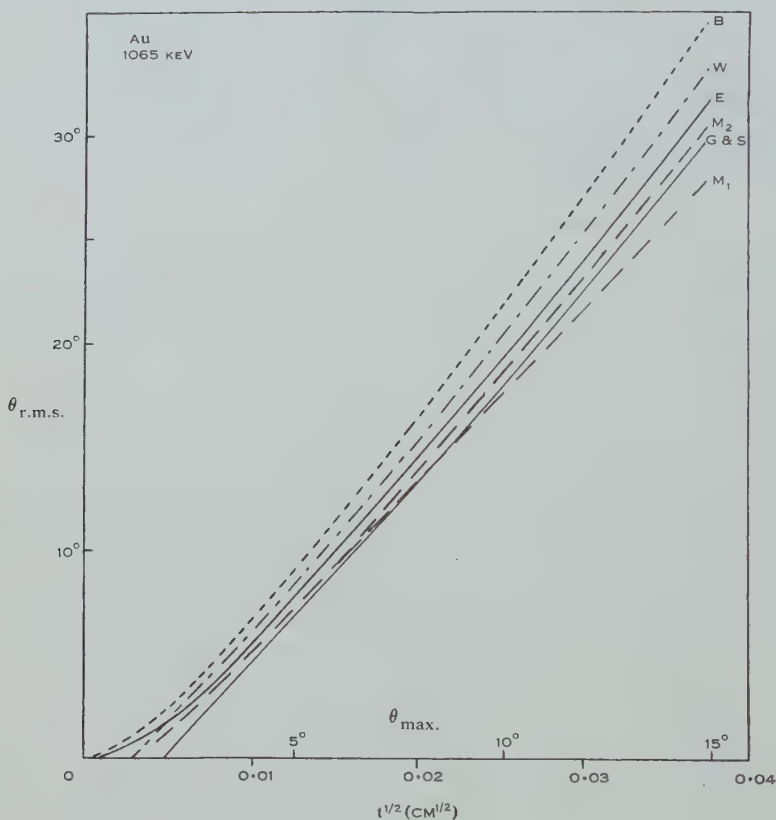


Fig. 2.—Multiple scattering of 1065 keV electrons in gold. Curves W , $G \& S$, and M_1 were obtained directly from the formulae for $\theta_{r.m.s.}$ given by Williams, Goudsmit and Saunderson, and Molière respectively (formulae (6), (8), and (10) of the present paper). Curves E , B , and M_2 were obtained by numerical integration of the single scattering distribution given by exact calculation, by the Born approximation with two-term field, and by Molière's formula (9.1) respectively (formulae (11), (16), and (17) of the present paper).

The Born and the Williams values of $\theta_{r.m.s.}$ lie above, and the Molière values below, the exact values. This is to be expected since the Born and Molière single scattering distributions lie respectively above and mostly below the exact distribution. Comparison of formulae (6) and (8) shows that the Goudsmit and Saunderson values of $\theta_{r.m.s.}$ lie below the Williams values by an absolute amount which increases with the thickness t , but by a fractional amount which decreases with increasing t .

The Molière curve M_2 lies closest to the exact curve as is to be expected, since it is based on the direct use of the single scattering distribution which lies closest to the exact distribution. The curve M_1 is based virtually on the use of the same single scattering distribution, and therefore coincides with curve M_2 over a fair range. The increasing separation of the curves M_1 and M_2 at layer thicknesses arises partly from the fact that a different value of $\theta_{\max.}$ is involved in the two cases. Thus one finds that at $t^{\frac{1}{2}}=0.025$, a change of only 1° in $\theta_{\max.}$ in the upper limit of the integral (4) alters $\theta_{r.m.s.}$ by 10 per cent. The agreement is good, however, up to $\theta_{r.m.s.}=20^\circ$, by which time the actual path length is appreciably greater than the thickness t , and energy loss is appreciable, so that the simple formulae no longer apply. The arbitrariness in the choice of $\theta_{\max.}$ cannot be avoided since some scatters will always occur at still larger angles, and in fact the distribution is not accurately Gaussian. The Molière theory gives the form of the complete distribution, with which comparison should be made in any experimental investigation.

For 121 keV electrons the various single scattering distributions in Figure 1 lie further apart than for 1065 keV electrons, and—as one would therefore expect—the values of $\theta_{r.m.s.}$ on the various theories show larger percentage differences amongst themselves than at 1065 keV. Also the exact curve for $\theta_{r.m.s.}$ is found to lie close to the Williams curve and well above the Molière curve. No practical significance, however, can be attached to results calculated for energies as low as 121 keV, since the thicknesses of foil for which values of $\theta_{r.m.s.}$ less than 30° occur are so small that too few collisions would occur for multiple scattering theory to be applicable.

The experiments of Oleson, Chao, and Crane (1941) at 6 MeV, and those of Kulchitsky and Latyshev (1942) at 2.25 MeV give, for heavy elements, values of $\theta_{r.m.s.}$ less than the Williams values by 10–15 per cent. Using (6) and (10) one finds that at these energies and for the values of $\theta_{r.m.s.}$ concerned, the Molière value of $\theta_{r.m.s.}$ is less than the Williams value by about the same percentage. This result indicates that, at these higher energies and for heavy elements, there is fairly good agreement with the Molière theory. This conclusion is also supported by the more recent experiments of Hanson *et al.* (1951) on the scattering of 15.7 MeV electrons in gold.

VI. MULTIPLE SCATTERING IN LIGHT AND INTERMEDIATE ELEMENTS

Let us now consider light and intermediate elements. Firstly there are the fairly recent experiments on beryllium, carried out by Hanson *et al.* (1951) with 15.7 MeV electrons. The observed values of $\theta_{r.m.s.}$ are less than Molière's by 3–7 per cent., and the discrepancy is thought to be possibly due to the use of the Fermi field, which would not be very accurate for as light an atom as beryllium.

The sensitivity of the scattering to the field was put to the test, not by the lengthy and tedious phase shift analysis, but by the approximate methods which should be fairly accurate for light elements at this high energy. In any case the methods should give accurately the difference in the scattering due to a small difference in the field.

As a first step the single scattering distribution was calculated in the following ways :

(a) using the Born approximation result (16) modified for a two-term field fitted to the Hartree field for beryllium,

(b) using the Born approximation result (16) modified for a three-term field fitted to the Thomas-Fermi field,

(c) using Molière's formula ((17) above), based on the WKB method and the Thomas-Fermi field.

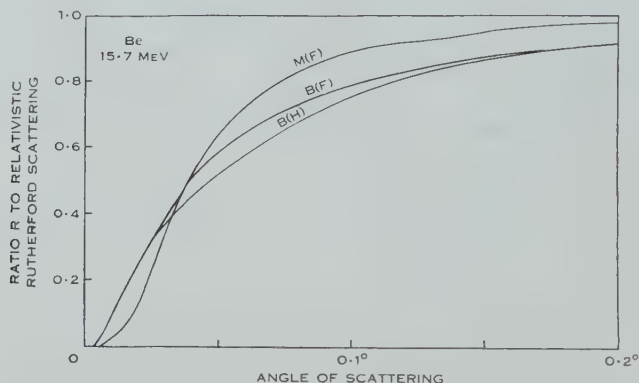


Fig. 3.—Angular distribution of single scattering of 15.7 MeV electrons in beryllium. Curves $B(F)$ and $B(H)$ were calculated using the Born approximation with the Fermi and Hartree fields respectively, and curve $M(F)$ using Molière's formula based on the WKB method and the Fermi field.

The three curves so obtained are shown in Figure 3. The differences between them are much less than the difference between the curves marked B_2 and M in Figure 1 for 1065 keV electrons in gold. One therefore expects fairly small differences in the values of $\theta_{r.m.s.}$ obtained from them; the more so since the effect of the Molière curve lying below the Born curves at the smallest angles is largely offset by its lying above at the larger angles.

These single scattering curves were then integrated numerically, using (4), to give values of $\theta_{r.m.s.}$ for the two thicknesses of foil used in the experiment. The results are given in Table 1.

Using in place of the Fermi field the more accurate Hartree field thus makes a difference in $\theta_{r.m.s.}$ of only $\frac{1}{2}$ per cent. with the Born approximation, and about the same difference should occur with the WKB method as used in the Molière theory. The discrepancy of 3–7 per cent. between theory and experiment can therefore hardly be attributed to the use of the less accurate Fermi field.

Experiments on carbon with 3–11 MeV electrons by Oleson, Chao, and Crane (1941) give values of $\theta_{r.m.s.}$ less than the Williams values by 10–15 per cent. About the same percentage difference is found between the Williams and Molière values, on substitution in (6) and (10). The results for carbon are therefore in fair agreement with the Molière theory.

Values obtained for elements of intermediate Z by Oleson, Chao, and Crane (1941) and by Kulchitsky and Latyshev (1942) are in fair agreement with the Goudsmit and Saunderson values, and substitution in (8) and (10) shows these to be higher by several per cent. than the Molière values. Taking into account the higher terms of the Molière multiple scattering distribution narrows it and merely increases the discrepancy. Theory thus lacks agreement with experiment over the whole range of Z .

TABLE 1

VALUES OF $\theta_{r.m.s.}$ FOR 15.7 MEV ELECTRONS IN BERYLLIUM, OBTAINED BY NUMERICAL INTEGRATION OF THE SINGLE SCATTERING DISTRIBUTION

Thickness (mg/cm ²)	Born Approxn. with Fermi Field	Born Approxn. with Hartree Field	WKB (Molière) with Fermi Field
257	2.18°	2.17°	2.14°
495	3.02°	3.01°	2.97°

The Molière theory is based on a more accurate treatment of the single scattering problem than the other theories, and might therefore be expected to give more nearly correct results for multiple scattering. Any discrepancies would be most likely for heavy elements, as suggested by the calculations in this paper; but the experiments indicate better agreement with Molière for heavy than for intermediate elements. Further experiments thus seem desirable.

VII. DIFFERENCE OF MEAN SQUARE ANGLE FOR ELECTRONS AND POSITRONS

There is a difference—which increases with Z —in the single scattering of electrons and positrons at large angles, due to electron spin; though the difference tends to zero as the angle tends to zero. One may therefore expect a difference in the value of $\theta_{r.m.s.}$ for electrons and positrons, though it is not obvious how large it will be. None of the multiple scattering theories gives a difference, and it must therefore be obtained by numerically evaluating the integral (4) for $\theta_{r.m.s.}$ using separate single scattering distributions for electrons and positrons.

(i) *Gold*.—Using the phase shifts given by Gunnarsen (1952) for 1070 keV positrons in mercury, the single scattering distribution was calculated and found to diverge appreciably from the corresponding curve for electrons as the angle increased from 5 to 10°. Then, using (4), it was found that, for foil thicknesses giving values of $\theta_{r.m.s.}$ less than 2°, the difference between $\theta_{r.m.s.}$ for electrons and positrons was less than 2 per cent.; as $\theta_{r.m.s.}$ increased to 10° the difference increased to about 6 per cent., where it remained for further increase in $\theta_{r.m.s.}$

Calculations were also carried out at 121 keV, making approximate estimates of the phases for positrons. It was found that for thicknesses giving a $\theta_{r.m.s.}$ greater than 10°, the difference between $\theta_{r.m.s.}$ for electrons and positrons was from 2–3 per cent.

The experimental results of McDonell (1953) on the scattering of 1 MeV electrons and positrons are consistent with a difference in $\theta_{r.m.s.}$, but its magnitude cannot be obtained accurately, since multiple scattering was involved only as a correction to single scattering observations at large angles.

(ii) *Argon*.—Estimates were made for argon, since Groetzinger, Humphrey, and Ribe (1951) report a difference of about 10 per cent. in the value of $\theta_{r.m.s.}$ for electrons and positrons over the range 0.3–2 MeV, the corresponding values of $\theta_{r.m.s.}$ varying from 10° to 3° .

For argon, $\alpha = Z/137$ is sufficiently small to allow one to use for small angles the approximate single scattering formula (McKinley and Feshbach 1948)

$$R = 1 - \beta^2 \sin^2 \frac{1}{2}\theta + \pi\alpha\beta \sin \frac{1}{2}\theta (1 - \sin \frac{1}{2}\theta), \dots\dots\dots (18)$$

with α positive for electrons and negative for positrons.

For an energy of 1 MeV, which is near the middle of the experimental range, the observed value of $4\frac{1}{2}^\circ$ for $\theta_{r.m.s.}$ is calculated from (5), (6), and (7) to correspond to $\theta_{max.} = 3^\circ$. From (18) the difference in R for electrons and positrons for an angle of 3° is 2 per cent. A similar calculation at the high and the low energy ends of the experimental range gives differences of 1 and 3 per cent. respectively in R . The value of the integrand in (4) at the upper limit of the integral differs for electrons and positrons by the percentages mentioned, and the differences will be less over the lower part of the range of integration, and hence less for the value of $\theta_{r.m.s.}$

Furthermore the formula (18) is for scattering by a Coulomb field, and at the angles in question screening will reduce the difference in R , and hence in $\theta_{r.m.s.}$, for electrons and positrons. The difference can hardly exceed 2 per cent. in any part of the experimental range of energies, but the observed difference is 10 per cent. Further experiments with other elements are clearly desirable in order to clarify the situation.

VIII. ACKNOWLEDGMENTS

The authors would like to thank Mr. J. A. McDonell for useful discussions. One of us (L.J.T.) was the holder of the Dafydd Lewis Scholarship while this work was carried out.

IX. REFERENCES

- BARTLETT, J. H., JR., and WATSON, R. E. (1940).—*Proc. Amer. Acad. Arts Sci.* **74** : 53.
 BETHE, H. A. (1953).—*Phys. Rev.* **89** : 1256.
 GOUDSMIT, S. A., and SAUNDERSON, J. L. (1940a).—*Phys. Rev.* **57** : 24.
 GOUDSMIT, S. A., and SAUNDERSON, J. L. (1940b).—*Phys. Rev.* **58** : 36.
 GROETZINGER, G., HUMPHREY, W., JR., and RIBE, F. L. (1951).—*Phys. Rev.* **85** : 78.
 GUNNERSSEN, E. M. (1952).—*Aust. J. Sci. Res. A* **5** : 258.
 HANSON, A. O., LANZL, L. H., LYMAN, E. M., and SCOTT, M. B. (1951).—*Phys. Rev.* **84** : 634.
 HOERNI, J. A., and IBERS, J. A. (1953).—*Phys. Rev.* **91** : 1182.
 KULCHITSKY, L., and LATYSHEV, G. (1942).—*Phys. Rev.* **61** : 260.
 McDONELL, J. A. (1953).—*Aust. J. Phys.* **6** : 245.
 MCKINLEY, W. A., JR., and FESHBACH, H. (1948).—*Phys. Rev.* **74** : 1759.

- MASSEY, H. S. W. (1942).—*Proc. Roy. Soc. A* **181** : 14.
- MOHR, C. B. O., and TASSIE, L. J. (1954).—*Proc. Phys. Soc.* (in press).
- MOLIÈRE, G. (1947).—*Z. Naturf.* **2a** : 133.
- MOLIÈRE, G. (1948).—*Z. Naturf.* **3a** : 78.
- MOTT, N. F., and MASSEY, H. S. W. (1949).—"The Theory of Atomic Collisions." 2nd Ed. p. 198. (Oxford Univ. Press.)
- OLESON, N. L., CHAO, K. T., and CRANE, H. R. (1941).—*Phys. Rev.* **60** : 378.
- ROSSER, J. B. (1951).—*J. Res. Nat. Bur. Stand.* **46** : 56.
- WATSON, G. N. (1948).—"A Treatise on the Theory of Bessel Functions." 2nd Ed. p. 385. (Cambridge Univ. Press.)
- WILLIAMS, E. J. (1938).—*Proc. Roy. Soc. A* **169** : 531.

RECENT STUDIES OF CHROMOSPHERIC SPICULES

By J. H. RUSH* and W. O. ROBERTS*

[Manuscript received January 15, 1954]

Summary

Five spicule film sequences taken on different dates in H α were studied. Maximum heights of spicules ranged downward from 20,000 km. Lifetimes ranged from the film limit of 0.3 min to about 30 min. Distributions of maximum heights and lifetimes indicated the occurrence of numerous small, short-lived spicules that were not resolved in these films. The mean upward velocity of the larger spicules was 31 km/sec, with a broad distribution. The absence of evidence of gravitational deceleration, together with a correlation between individual velocities and maximum heights above the chromosphere, suggests a modified interpretation of the spicule process. The data yield a tentative upper limit to the number of spicules on the entire Sun of 2.2×10^4 .

I. INTRODUCTION

During the past 10 years, chromospheric spicules have gained steadily in interest and significance. They are involved in theories of the structure and conditions in the chromosphere, and of the transport of material and kinetic energy into the corona; they are probably related to the granular turbulence structure of the photosphere; and they even suggest a connexion with the origins of M-region storms.

Observers of eclipses have noted the profusion of spikes and filaments that emanate from the upper chromosphere. On the basis of such observations, Menzel (1931) suggested that the chromosphere may be composed entirely of such structures. Roberts's discovery in 1943 that the spicules could be observed with the Lyot-type coronagraph attracted renewed interest to the subject. Roberts's earlier papers (Roberts 1945; Roberts *et al.* 1949) on spicules have reported the results of preliminary studies.

Hedeman (1949) has reported a study of very small prominences within the size range of spicules. Mohler (1951) obtained heights and numbers of spicules from eclipse observations. Dizer (1952) has reported an investigation of equatorial spicules; he finds systematic differences in the behaviour of spicules associated with quiet or agitated regions of the chromosphere.

Thomas (1948) called attention to the significance of spicules to the theoretical interpretation of the chromosphere. His theory sought to explain the broadened line profiles and the high temperatures found for the chromosphere by Redman (1942), which agreed with the density gradients of Wildt (1947), in terms of *mechanical* transport of energy from the photosphere to the

* High Altitude Observatory of Harvard University and University of Colorado, Boulder, Colorado, U.S.A.

atmosphere of the Sun. He developed a theoretical mechanism of such transport in the form of supersonic jets, which he assumed correspond to the spicules. Recently doubts were raised by Woolley and Allen (1950) as to the existence of a high chromospheric temperature. Redman (1953), however, found from 1952 eclipse data a new value of $17,000^{\circ}\text{K}$. The evidence for a high temperature has been reinforced by the recent findings by Athay *et al.* (1953), Dimock *et al.* (1953), and others from High Altitude Observatory spectra of the 1952 eclipse. From studies of free-bound emission in the Balmer continuum, they have obtained a strong height gradient of temperature, leading to values of the order of $25,000^{\circ}\text{K}$ or more in the upper chromosphere.

The identification of the photospheric granules with turbulent cells in the convection layer inevitably suggests the possibility that the spicules originate in this turbulence and may even be projections through the chromosphere of individual granule cells. We believe that the research reported here strengthens the hypothesis of such a direct relation between spicules and granules, though it by no means disposes of all the difficulties. The bases for this statement will be developed in detail in our conclusions.

Allen (1944) suggested that the white-light coronal streamers, rotating with the Sun, are the agents of the periodic M-region storms. The authors and others (Roberts, Grenchik, and Billings 1953; Rush and Roberts 1953), following up this suggestion, have proposed the hypothesis that such coronal streamers are neutral beams of ionized material ejected in the spicules and focused by the interactions of the magnetic fields of sunspot groups with each other, or with a general solar field.

We undertook the present research to develop as detailed as possible an understanding of the behaviour of spicules from available observations, and particularly to evaluate more accurately their numbers, dimensions, and velocities—the parameters that bear most significantly on the questions of material and energy transport. Our results generally confirm those already reported. Certain inconsistencies with other work are noted.

II. OBSERVATIONAL MATERIAL

Five 35-mm film sequences exhibiting spicule activity were studied. Table 1 lists the pertinent data for these films. All were exposed on Eastman 103-Hz film in the 5-in. coronagraph at Climax, through a birefringent filter having a half-width of approximately 4 \AA , centred on H α . About 90° of the limb was included in the picture frame. Exposure times were approximately 2 sec.

These five films are the best that were obtained before the end of the detailed reduction program. Preliminary studies of films Nos. 1 and 3 have been reported by Roberts *et al.* (1949). Nos. 4 and 5 are distinctly superior to the first three, because of their higher time-resolution and much richer spicule activity. Whether the latter effect is due to better seeing or actually greater spicule abundance is uncertain.

Reductions of these films were carried out by a computations group at the Massachusetts Institute of Technology. The quantitative data were derived from measurements on a Mann comparator of spicule positions and displacements.

III. RESULTS

(a) General Remarks

Because of the richness, homogeneity, and superior time resolution of films Nos. 4 and 5, we have drawn our conclusions largely from them. Several factors tended to introduce spurious latitude dependences into the spicule distributions. We do not exclude the possibility of real correlations between spicule characteristics and latitude; but the present data have not afforded means of distinguishing any such real dependence from the spurious effects that we know are present. We have therefore assumed that the fairly homogeneous characteristics of the spicules within the middle segment of about 50° of the limb in our pictures are an acceptable sample of phenomena which provisionally we assume to be uniformly distributed over the entire Sun—at least in regions free of sunspot activity. The results that follow are derived from this central segment of the pictures.

TABLE 1
SUMMARY OF SPICULE FILMS

Film No.	Date	Latitude at Middle of Picture (+ = North)	Duration (min)	Interval between Exposures (sec)	No. of Spicules Observed
1	12.xii.43	+78°	235	60	361
2	28.xii.43	+86°	67	20	126
3	21. ii.46	+84°	122	20	207
4	29. i.49	—86°	50	10	458
5	2.vii.49	—88°	42	10	443

We wish to emphasize that the spicules observed for this study are faint objects, compared with the usual quiescent prominence. They completely surmount the chromosphere proper. With the exposures used in this work, the chromosphere itself appears as an overexposed luminous band with a rather irregular upper edge. Many of the photographs of spicules taken by others and compared with ours appear to have been taken at somewhat shorter exposures, and thus refer to levels probably lower in the chromosphere.

(b) Time and Space Distribution of Spicules

Analysis of the position angles and beginning times of the spicules of film No. 4 showed no significant tendency to occur in non-random groups either in time or in space, except the dubious latitude distribution already mentioned. A similar analysis of the distribution of the non-radial spicules only, in film No. 5, showed no non-random tendency.

The density and uniformity of distribution of spicules over the Sun are of fundamental interest. The five films that have been reduced in detail cover only the polar zones; but other films in our files show spicules at low latitudes. Hansen and other observers at Climax (unpublished data) have reported occasions when the entire limb in $H\alpha$ was "bristling" with spicules. Mohler (1951) has

reported on the profusion of spicules appearing in white light from pole to equator in Marriott's 1930 eclipse photographs. Dizer (1952) has reported a study of equatorial spicules in Lyot's films. R. B. Dunn has communicated unpublished observations of spicules over wide ranges of latitude. Yet we have seen quiet regions of the limb in $H\alpha$ photographs that were entirely free of spicules, even though the definition of the limb and adjacent prominence detail appeared adequate for the detection of spicules; and Roberts is of the opinion, based on visual observation, that spicules tend to be most uniform in behaviour—simple, radial, and undisturbed-looking—in large regions centred roughly on the poles.

On the basis of all available evidence, we believe it is reasonable to conclude that spicules occur at one time or another over the entire chromosphere; but the available evidence does not seem sufficient to establish whether or not they are active *continuously* over the entire Sun.

In examining some good copies of Swarthmore College and Lick Observatory eclipse plates,* we noted that in several instances spicules appeared superimposed on the white-light coronal arches. In each such case the spicules departed from the radial direction and were aligned tangent to the contours of the arches.

If it be assumed that the distribution of spicules we observed in the vicinity of the pole is uniform over the entire Sun, then an upper limit to the number of spicules on the Sun at a given time can be determined, as Mohler (1951) has demonstrated. The method is to assign to each spicule an area of the chromosphere equal to the square of the mean distance observed between spicules at the limb. In each of our films Nos. 4 and 5, an average of 42 spicules per radian were on the limb in the region of interest at a given time. The result, on this basis, for the entire Sun is 2.2×10^4 spicules.

An independent check of spicule numbers was made on the flash spectrum of the 1952 eclipse. Spicules appeared clearly in $H\alpha$ at a height of 6300 km above the photosphere, over an arc of the east limb centred at about latitude 65° . The spicule count over the middle 45° of this arc was 46 per radian.

This procedure assumes that all of the observed spicules are actually on the limb. Since they must be distributed in the line of sight, the actual number on the Sun is substantially less than the above result. A rough estimate of the actual distribution, based on the argument developed later (Section III (d)), suggests that 5000 is a reasonable number. In appraising this result, it must be kept in mind that the visibility of small spicules is very sensitive to seeing conditions and other factors.

(c) *Size and Brightness of Spicules*

The reduction programme included a qualitative estimate of the relative size and brightness of each spicule. The size so estimated was a composite of maximum height and breadth, an index of the apparent bulk of the spicule. No significant correlation was found between size and brightness.

* Through the courtesy of Dr. Donald H. Menzel of Harvard College Observatory.

The actual diameters of spicules are quite uncertain because of their diffuse boundaries and the limited resolution in the photographs. For the great majority of spicules, a typical diameter at the chromosphere of the order of 2000 km is indicated.

(d) Heights of Spicules

The maximum heights reached by spicules were difficult to determine, because of the faintness of the tip of the spicule, uncertainty in locating the level of the chromosphere, and inconsistencies caused by variations in seeing. Because of such difficulties, it was impracticable to determine the heights of the great majority of spicules, those less than two or three thousand km in height above the photosphere.

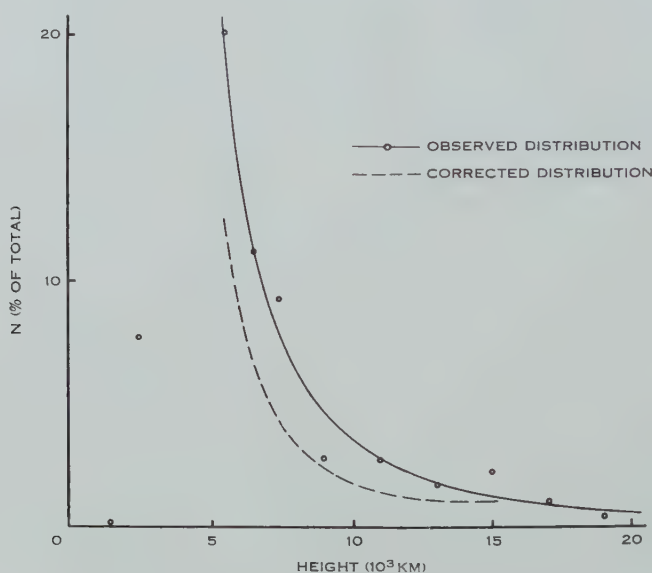


Fig. 1.—Distribution of heights above the chromosphere of 263 spicules.

The distribution of observed heights from films Nos. 4 and 5 is shown in Figure 1. Probably the distribution is influenced significantly by observational selection in the reduction process; the heights of only one-fourth of the spicules were measured, and the decision whether to try to measure a spicule or omit it was necessarily somewhat arbitrary. It appears virtually certain that the decline of the curve below 5000 km is due to such selection, and that the distribution actually continues upward somewhat as indicated in the figure. This distribution is similar to that of lifetimes (Section III (e)); but the data were insufficient to determine its form with comparable precision.

The distribution of spicules in the line of sight obviously affects their apparent heights, since they suffer partial obscuration depending on their distances from the limb. If a spicule of actual height h is located at an angle α

from the limb, its apparent height h^* is given with sufficient accuracy by the approximation

$$h^* = h - \frac{R\alpha^2}{2},$$

where R is the radius of the limb. Thus the apparent heights h^* of a set of spicules of a given actual height h are dispersed throughout the range from 0 to h .

The actual height distribution $N(h)$ may be approximated from the observed distribution $N(h^*)$ by a simple geometrical analysis (Appendix I). The dashed curve in Figure 1 is the corrected distribution we obtained in this way from the solid curve representing the observed distribution. The value 15,000 km was taken as the maximum height H , and computations were based on 1000-km ranges from 5000 to H km. Evidently the effect of distribution in the line of sight does not greatly alter the form of the distribution, which after correction still approximates a hyperbola or an exponential.

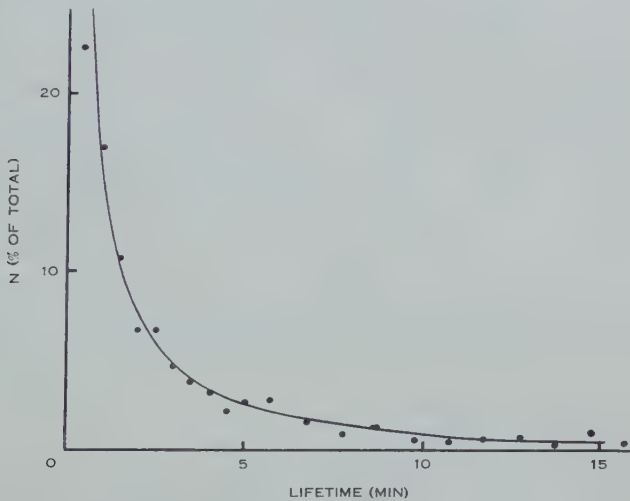


Fig. 2.—Distribution of lifetimes of 400 spicules from films Nos. 4 and 5. Curve: $N\% = 16.6/T - 0.7$.

(e) *Lifetimes of Spicules*

We define the lifetime T of a spicule as the time interval from its first appearance to its final disappearance. During this period a typical spicule rises vertically at roughly constant velocity from the chromosphere to its maximum height; then it may remain stationary while fading from visibility, or it may appear to descend back into the chromosphere at a velocity comparable to the upward velocities.

The frequency distribution of lifetimes for 400 spicules is given in Figure 2. The mean lifetime for the range covered in the figure is 3.3 min. Similar distributions were obtained from films Nos. 1, 2, and 3, yielding an overall mean lifetime for all five films of about 3.5 min. These means agree well with Roberts's

(1945) preliminary result of 4–5 min from film No. 1, which included no lifetimes less than 1 min.

The hyperbola ($N\% = 16.6/T - 0.7$) approximates the data much more closely than does an exponential. The continuing upward trend of the curve at the short-lifetime end indicates that the maximum is to be found below our present 0.3-min limit of time resolution, and that large numbers of short-lived spicules must occur below this limit. This result lends some support to the view that the chromosphere is composed entirely of radial spikes or filaments.

The evaluation of T is complicated by two important factors. First, the spicules seen in profile on the limb presumably are distributed randomly in the line of sight; and no simple theoretical correction to the distribution for obscuration of spicules by the limb can be applied, because T is a composite of the interval during which the spicule is ascending (or descending) and the interval during which it is stationary. The former interval is affected by limb obscuration, but the latter is not, and the two do not appear to be systematically related.

As we showed in Section III (*d*), however, the effect of obscuration by the limb is to elevate the height distribution curve by a factor that, qualitatively speaking, varies inversely as the height. Only a weak correlation exists between heights and total lifetimes of the spicules whose heights were measured; but there is good reason to believe that the correlation would be stronger if the heights of all spicules could be measured. The heights of most spicules were too small for reliable measurement, and these were predominantly in the short-lifetime region. If we assume such a rough overall correlation between spicule heights and lifetimes, it is evident that obscuration by the limb of spicules distributed along the line of sight will tend to elevate the lifetime distribution curve in the same sense as the height curve, so that the short-lived spicules appear relatively too numerous. Although this effect cannot be accurately evaluated from our data, it is evident that its influence on the distribution of lifetimes must be less than on that of heights, and that the effect cannot alter the general form of the distribution.

Second, intervals of bad seeing and possibly other factors prevent a spicule from being seen on every frame during its lifetime. Only continuity of position and development permits the identification of such an interrupted sequence as a single spicule. Obviously, such interruptions must be particularly effective in cutting off observation of a spicule at the beginning and end of its lifetime, when it is least conspicuous. This loss of terminal frames results in a systematic reduction of the observed lifetimes below the true values.

In this case, the effect on the distribution curve can be estimated from the number of frames in which a spicule was not seen, between its first and last appearance. The resulting corrections to the observed distribution are small, amounting to about 8 per cent. for the most numerous lifetime class. But this effect of lost terminal frames upon the actual lifetime distribution curve is in the opposite sense to that of obscuration in the line of sight, because it preferentially lowers the short-lifetime end of the curve. Lacking a way to evaluate the line-of-sight effect, we can only note that these two effects tend to compensate. We believe that the empirical distribution curve of Figure 2 is a good

approximation to the actual distribution of spicule lifetimes ; and, in particular, that the maximum of the distribution is to be found at shorter lifetimes than we have resolved. The relatively low value of the last point ($T=0.5$ min) may mean that we are approaching the maximum ; but it may equally well be ascribed to failure to detect some of the very short-lived spicules.

(f) *Velocities of Spicules*

The velocities of some larger spicules were determined from plots of the heights measured on successive frames. The results are summarized in Table 2. In this summary, only those spicules have been included whose development was fairly regular and continuous. Such a spicule typically rose with approximately constant velocity and stopped abruptly at its maximum height. In

TABLE 2
SUMMARY OF DATA ON ASCENDING VELOCITIES (v_a) OF SPICULES

Film No.	No. of Typical Spicules	Mean Lifetime* (min)	v_a (km/sec)		
			Mean	Median	Mode
1	17	11.3	33	30	—
2	6	7.2	30	—	—
3	15	15.0	29	32	34
4	35	10.5	30	26	28
5	28	10.8	33	24	25
All	101	11.1	31	—	—

* Mean lifetime of those spicules whose entire lifetimes were observed. Note that these lifetimes for spicules with measured velocities are substantially longer than the mean lifetimes of all spicules on these films.

view of the many uncertainties and irregularities involved, the agreement among the mean velocities derived from the five films is remarkably close. These values and the mean of all of the velocities, 31 km/sec, confirm the value of 30 km/sec estimated by Roberts (1945) in a preliminary study of film No. 1. They agree also with the mean of Dizer's (1952) values of 20 and 40 km/sec for equatorial spicules in "quiet" and "agitated" regions of the chromosphere.

It should be noted, however, that the values listed in Table 2 are the means of a broad range of individual velocities. The distribution of velocities derived from films Nos. 4 and 5 is presented in Figure 3.

Some spicules exhibited apparent variations in velocity, occasionally reaching 200 km/sec, and sometimes stationary phases, before attaining their maximum height. In a few cases spicules seemed to ascend in a succession of surges. We did not attempt to summarize the velocities with which some spicules appeared to descend. These descents sometimes were regular and measurable ; more often they were abrupt and irregular, suggesting that an apparent descent was actually a progressive loss of visibility. It might be

thought that variations in seeing from frame to frame would affect the apparent height of a spicule, thus introducing spurious irregularities into its progress; but we found a systematic variation of only about 10 per cent. in height between the best and worst seeing.

The generality of the results in Table 2 is limited also by the fact that satisfactory measurements of velocity could be made only on relatively large, long-lived spicules. Almost all of these exceeded 5000 km in height, and their mean lifetime of 10 min contrasts with a mean of about 3.5 min for all spicules that were observed.

The individual velocities showed no correlation with lifetimes. They were, however, correlated with maximum heights by a coefficient of 0.70. This relation (Fig. 4) is essentially linear, with its origin approximately at the top of the chromosphere. These relations appear to contradict those reported by Dizer (1952), who found uniform speeds of ascent and descent, and heights directly proportional to lifetimes. The scatter of the data is such that a curve originating at the photosphere is not entirely precluded: but such a curve, to fit the data acceptably, would have to be non-linear.

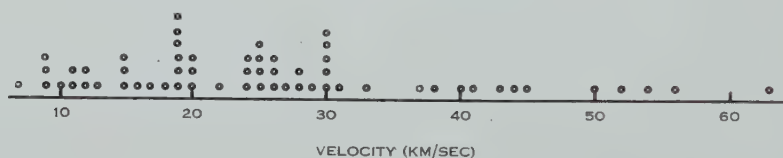


Fig. 3.—Distribution of ascending velocities of 59 spicules from films Nos. 4 and 5.

It is obvious that the correlation between heights and velocities, implying that these spicules must rise to maximum height in about 4.5 min, cannot be generalized to the large class of spicules whose total lifetimes are less than 4 min. Yet we found no basis for attributing the correlation to any preferential selection of data. Though the velocity measurements were limited of necessity to a class of generally large, long-lived spicules, the range of heights, velocities, and total lifetimes within this body of data was great.

Figure 4 shows no evidence of influence of line-of-sight obscuration on the height distribution. The scatter of heights about the visually-located line is essentially random; the points show no tendency to concentrate in the upper portion of the plot, as they would be expected to do if obscuration in the line of sight were a dominant factor. It seems probable that the necessity of selecting relatively large spicules for velocity measurements insured that most of those selected would be situated near the limb, thus practically eliminating the complications that would have resulted from an extensive distribution in the line of sight.

IV. THEORETICAL IMPLICATIONS

One of the most interesting of our results is the finding that the linear relation between velocity and height apparently converges to an origin approximately at the top of the chromosphere. As it does not appear probable that the

jet, which emerges as a spicule, originates at so high a level, the simplest conclusion is that while the spicule is rising through the chromosphere it remains in condition to radiate effectively in $H\alpha$; it might be regarded as a moving sample of the chromosphere. But, as it rises above the chromosphere, it enters a region in which, because of increasing ionization, the radiation from neutral hydrogen is rapidly attenuated. The radiation from the rising material should, however,

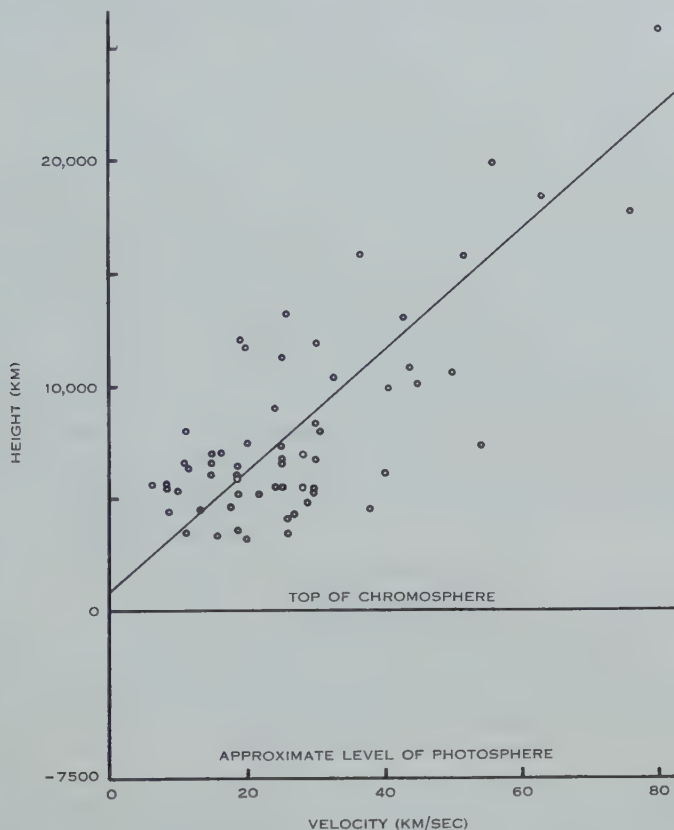


Fig. 4.—Ascending velocities and maximum heights of 61 spicules from films Nos. 4 and 5. Slope of curve: $\Delta h/\Delta v = 4.5$ min.

require some time to decline to a negligible level, and the height above the chromosphere to which it remains visible should then be proportional to its velocity, to a first approximation.

On this interpretation, the observation that a spicule typically rises at approximately uniform velocity and then abruptly stops becomes intelligible. The interpretation we propose is that the material continues to flow upward beyond its visible terminus at a constant velocity during the entire period in which the spicule remains apparently stationary. Then, if the velocity decreases, the column of material appears to descend; or, if the supply of material diminishes, the spicule fades away without appearing to descend. The anomalous cases of spicules that appear to rise and fall repeatedly in an irregular series of

surges can be explained by assuming corresponding fluctuations in density of material or in upward velocity alone.

This interpretation of a spicule implies that the motion of the jet is essentially independent of the gravitational field. Our measurements of spicule heights in successive frames indicated that the upward velocity of a typical spicule is essentially constant. The fluctuations appeared to be random, with no evidence of systematic deceleration. It appears that spicules, like eruptive prominences on a larger scale, ascend at approximately constant velocities despite the gravitational field.

Thomas (1948) estimated that the kinetic energy transported by the spicules from the photosphere is approximately 2×10^{32} ergs/sec, of which about 10 per cent. is dissipated in the chromosphere. This estimate was based on the assumption of gravitational motion of the spicule material, which implied an initial velocity of about 90 km/sec at the photosphere to project a spicule to a typical height of 7500 km above the chromosphere. Our findings, however, indicate that the velocities at the photosphere must be substantially the same as those observed at the top of the chromosphere. Our mean velocity of 31 km/sec therefore requires a reduction of Thomas's estimate of the kinetic energy in the spicule system by a factor of 9; but, on our interpretation, all of this energy would be available for heating the chromosphere and corona.

We do not believe, however, that our mean of 31 km/sec for the few velocities we measured is representative of all spicules. The relation between height and velocity, if it has any validity for the numerous unmeasured small spicules, implies that these must have velocities of less than 10 km/sec. At the same time, the distribution of lifetimes indicates the existence of many still shorter-lived small spicules not yet observed individually. The low energies of these small jets may be effectively compensated by their great numbers. We do not believe that the new data are necessarily inconsistent with Thomas's estimate of energy transport; but they raise questions that can be resolved only by further research.

The possibility of a direct relation to the photospheric turbulence granules is implicit in any discussion of spicules. The chief obstacle to direct identification of the two phenomena is the wide disparity between the mean turbulence velocities observed in the granulation and the observed velocities of the larger spicules. Yet the relation we have cited between heights and velocities, as well as the scarcity of spicules having velocities great enough to measure and the large scatter of velocities in this group, all suggest a broad statistical distribution of spicule velocities, only a few of which are as great as 30 or even 10 km/sec. Similarly, the mean Doppler velocity of the order of 1 km/sec observed in the granulation by Richardson and Schwarzschild (1950) and others does not preclude a statistical dispersion of velocities such that perhaps one granule in 10,000 might exceed 10 km/sec.

Nor does the disparity between the observed numbers of spicules and of granules appear to offer any insuperable difficulty. We found a value of the order of 10^4 for the total number of observable spicules on the Sun, under certain

dubious assumptions; Mohler (1951), from eclipse plates of extraordinarily good definition, found an upper limit of 4×10^5 . The number of small granules, from observations by Keenan (1938, 1939), Macris (1953), W. A. Miller (personal communication), and others, appears to be about 10^6 . Both the "grasslike" appearance of the upper chromosphere and the trend of the lifetimes distribution argue for the existence of additional small unresolved spicules which may quite possibly be more numerous than those we have observed by a factor of 100.

We believe, therefore, that the evidence at hand supports the hypothesis that the quiet-Sun chromosphere—or at least its visible components—is composed of radial jets arising from the photospheric turbulence. On this view, each convection cell of ascending gas, marked by a bright granule on the photosphere, projects a jet beyond the photosphere. The velocities of these jets, derived from the underlying turbulence, are distributed statistically over a range of perhaps 0–100 km/sec. The uneven top of the chromosphere is defined statistically as the region in which the great majority of small, low-velocity jets encounter conditions that render them invisible. The positive temperature gradient in the chromosphere is maintained, as Thomas (1948) proposed, by the dissipation of kinetic energy from the rising jets into the surrounding gas. The jets of exceptionally great bulk and velocity, the "tail" of the statistical distribution, remain visible to various heights above the general level of the majority, and appear individually as spicules.

This hypothesis is admittedly tentative and subject to some unresolved difficulties; but we believe that it fits the observational situation sufficiently well to merit intensive study.

V. ACKNOWLEDGMENTS

The authors would like to express their appreciation of the contributions that have made this research possible. The observational data were derived from the Climax observing programme, supported in part by the National Bureau of Standards. The reduction measurements were carried out by a group at Massachusetts Institute of Technology under the direction of Dr. Zdenek Kopal assisted principally by Miss Virginia Brenton and Mrs. Martha B. Shapley. This reduction programme was supported in part by the Office of Naval Research under Contract N8onr-64801, and in part by the Air Materiel Command under Contract W19-122ac-17. Some preliminary analyses of the reduced data were made by Mr. B. Witte, then of the High Altitude Observatory staff, with the support of the Research Corporation. The present study has been supported principally by the Radio Corporation of America under a research contract with the High Altitude Observatory.

VI. REFERENCES

- ALLEN, C. W. (1944).—*Mon. Not. R. Astr. Soc.* **104**: 13.
ATHAY, R. G., BILLINGS, D. E., EVANS, J. W., and ROBERTS, W. O. (1953).—*Astr. J.* **58**: 210.
DIMOCK, D. L., BILLINGS, D. E., ATHAY, R. G., and THOMAS, R. N. (1953).—Paper presented before the American Astronomical Society in Boulder, Colorado, August 29, 1953. Abstract, *Astr. J.* **58**: 213.
DIZER, M. (1952).—*C.R. Acad. Sci., Paris*, **235**: 1016.

- HEDEMAN, E. R. (1949).—*Publ. Astr. Soc. Pacif.* **61** : 224.
 KEENAN, P. C. (1938).—*Astrophys. J.* **88** : 360.
 KEENAN, P. C. (1939).—*Astrophys. J.* **89** : 604.
 MACRIS, C. (1953).—*Ann. Astrophys.* **16** : 19.
 MENZEL, D. H. (1931).—*Publ. Lick Obs.* **17** : 287.
 MOHLER, O. (1951).—*Mon. Not. R. Astr. Soc.* **111** : 630.
 REDMAN, R. O. (1942).—*Mon. Not. R. Astr. Soc.* **102** : 140.
 REDMAN, R. O. (1953).—*Accad. Naz. Lincei, Conv. Volta No.* 11.
 RICHARDSON, R. S., and SCHWARZSCHILD, M. (1950).—*Astrophys. J.* **111** : 351.
 ROBERTS, W. O. (1945).—*Astrophys. J.* **101** : 136.
 ROBERTS, W. O., BRENTON, V., SHAPLEY, M. B., and KOPAL, S. (1949).—*Publ. Astr. Soc. Pacif.* **61** : 160.
 ROBERTS, W. O., GRECHIK, R., and BILLINGS, D. E. (1953).—Paper presented before the American Astronomical Society in Boulder, Colorado, August 29, 1953. *Abstract, Astr. J.* **58** : 225.
 RUSH, J. H., and ROBERTS, W. O. (1953).—*Trans. Inst. Radio Engrs., Professional Group on Communications Systems CS-2* : 24.
 THOMAS, R. N. (1948).—*Astrophys. J.* **108** : 130.
 WILDT, R. (1947).—*Astrophys. J.* **105** : 36.
 WOOLLEY, R. V. D. R., and ALLEN, C. W. (1950).—*Mon. Not. R. Astr. Soc.* **110** : 358.

APPENDIX I

Effect of Obscuration in the Line of Sight

In Figure 5, a, b, c, d, e represent an arbitrary series of equal ranges Δh of actual spicule heights, from the limb L to the maximum height H above the limb. The angles A, B, C, D, E are the successive angular increments from the limb over which the curvature of the limb away from the tangent L is equal to $\Delta h/2, 3\Delta h/2, 5\Delta h/2, 7\Delta h/2, 9\Delta h/2$ respectively. These angles are small enough that the reduction of h by foreshortening can be neglected. Also, the ratio H/R is sufficiently small that the curvature away from the tangent over any angle we are concerned with can be considered constant for all radii from R to $R+H$.

An observer whose line of sight is parallel to the tangent to the limb L will see spicules apparently in range a arising throughout the angular range $A+B+C+D+E$; but it is evident from the figure that this apparent class of heights is actually a composite derived from all values of h . Heights actually in range a are observed in this range over the angle A ; heights actually in range b are observed as being in range a throughout angle B ; those actually in c appear in a throughout angle C ; and so on. Thus the heights appearing in range a actually are derived from all height ranges. The apparent range b , however, receives no contribution in angle E ; c receives none in angle D or E ; and so on. It is therefore evident that, if spicule heights are randomly distributed over the surface of the Sun, the distribution observed in the line of sight will be a distortion of the actual distribution, and will show disproportionately large numbers of spicules in the lesser height ranges.

By examination of the observed distribution curve $N(h^*)$, a somewhat arbitrary determination may be made of the maximum height that occurs with any significant frequency. This height is taken as the maximum actual height

H attained by the spicules ; but the correction is not very sensitive to the choice of H , because of the relatively small numbers of spicules in this part of the distribution. The angles A, B, C, D, E are then computed within which the mean height of range e falls in the arbitrary height ranges e, d, c, b, a respectively. Since the maximum range e can receive no apparent contributions from the lesser ranges, all of the N_e^* spicules appearing in range e are assumed to be located

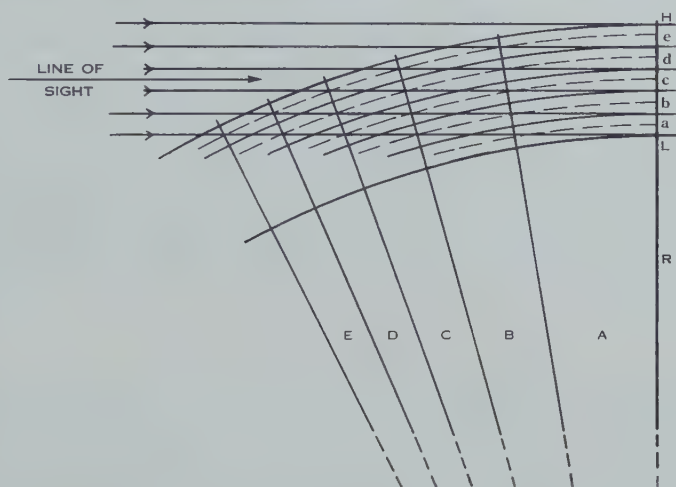


Fig. 5.—Effect of obscuration by the limb on apparent heights of spicules distributed along the line of sight.

in the angle A , so that $N_e^* = N_e$. Now, the N_d^* spicules appearing in range d include $(B/A)N_e$ spicules from actual range e distributed over angle B ; and the difference between this term and N_d^* gives the actual residue N_d in angle A . Similarly, $N_c = N_c^* - (B/A)N_d - (C/A)N_e$; and so on. The numbers N_a, N_b, \dots of spicules occurring over the same angular range A therefore constitute the actual height distribution.

ON THE CHOICE OF GLASSES FOR CEMENTED ACHROMATIC APLANATIC DOUBLETS

By W. H. STEEL*

[*Manuscript received November 30, 1953*]

Summary

It is shown that the types of glass given by tables for thin, low aperture telescope objectives of the "crown leading" doublet form are also those most suitable for thick, high aperture systems. The maximum focal lengths and the tolerances permissible in glass properties are given for each aperture, and the variation of these properties is considered when the thickness or wavelength of correction is changed, or the objective used with prisms.

I. INTRODUCTION

Three variables are required to design a cemented doublet telescope objective of the Clairaut type, corrected simultaneously for longitudinal chromatic aberration, spherical aberration, and coma. If the focal length is fixed, two of these variables are provided by the relative powers of the crown and flint components and the bending of the lens as a whole, and the third must be chosen from the indices and dispersions of the two glasses used. Existing glass types do not provide a continuous variation of these properties, and hence some means is required of choosing suitable glass pairs.

Several investigators have given methods for choosing suitable glasses for thin, low aperture objectives. Harting (1898) gives tables for this purpose, and Moffitt (1925) and Moffitt and Kaspereit (1925) give a graphical method. Very complete tables based on third order aberration theory are given by Brown and Smith (1946).

This paper discusses the glass properties required for objectives of large relative apertures for which higher order aberrations become appreciable and thick components must be used. The "crown leading" case only is considered, as this has the smaller contact curvature.

A series of doublets was designed at angular semi-apertures from 0 up to 0.15 by trigonometrical methods: a pair of indices was chosen for the crown and flint components and the spherical aberration and coma corrected by ray traces. The most suitable form of the lens thus having been found for each aperture, the relative dispersions of the two glasses necessary for chromatic correction were calculated. The residual secondary aberrations were also calculated and expressed as the maximum scale (maximum focal length) to which the objective could be made without exceeding aberration tolerances.

* Division of Physics, C.S.I.R.O., University Grounds, Sydney.

Two pairs of glasses were initially studied in detail, chosen to represent the two extremes of refractive index difference within the range of glasses commonly used for doublets. It was found that the variations of relative dispersion and maximum focal length with aperture were practically the same in both cases, and further calculations with an intermediate pair of glasses at an angular semi-aperture of 0.15 confirmed this observation. Hence it may be assumed that the results of this investigation should apply to any glass pair whose properties lie between those of the two pairs investigated.

As variations of the thickness of each component could disturb the results, the thicknesses were chosen to be a constant fraction of the lens aperture. However, it was found that quite considerable variations from this standard thickness had very little effect on the choice of glass. Other effects considered were a change of the wavelength at which the spherical aberration was corrected, the results of under- or over-correction of the aberrations, and the tolerable variations of the relative dispersions at each aperture.

II. METHOD OF CALCULATION

The Brown and Smith (1946) tables give, in terms of the indices of the two glasses, the curvatures of the surfaces and the required relation between the glass dispersions for chromatic correction. This is given as $N = (v_a + v_b) / (v_a - v_b)$, where v_a and v_b are the Abbe values for the crown and flint components respectively. In this paper it is found more convenient to use the variable

$$\rho = \frac{v_a}{v_b} = \frac{N+1}{N-1}.$$

For lenses of finite aperture, the third order aberrations should not be corrected completely but should be used to balance the aberrations of higher orders. If these higher order aberrations are assumed to be predominantly of the fifth order the best image is obtained when the ray aberrations are corrected at the apertures given by Maréchal (1947), namely, the spherical aberration at the margin and the coma at $\sqrt{1.2}$ times the marginal aperture. It was found to be more convenient to trace an extra ray at this $\sqrt{1.2}$ aperture than to balance the marginal and primary coma using the rays traced to find the spherical aberration. For each pair of indices chosen, the coma was made zero by bending the lens, and then the relative powers of the two components were adjusted until the spherical aberration was also zero.

To correct the lens for longitudinal chromatic aberration, the ratio ρ of the two v values was calculated by equating to zero Conrady's (1904) formula for chromatic path differences

$$\Sigma E = \Sigma (d' - D') \delta n, \quad \dots \dots \dots (1)$$

where d' and D' are the thicknesses of each component along the axis and along the ray and δn is the index difference between the wavelengths at which the

primary aberration is corrected. If the C and F rays are brought to the same focus, it follows that

$$\rho = -\frac{(n_d-1)_a(d'-D')_a}{(n_d-1)_b(d'-D')_b}, \quad \dots\dots\dots (2)$$

where the suffices *a* and *b* refer to the crown and flint components respectively.

This equation was used as recommended by Conrady for a marginal ray. However, for angular semi-apertures greater than 0.1, the longitudinal chromatic aberration was found to be corrected somewhat inside the $1/\sqrt{2}$ zone. If equation (2) is applied instead to the ray at the $\sqrt{1.2}$ aperture used to calculate the coma, the aperture at which the chromatic aberration is zero is very close to the $1/\sqrt{2}$ zone, but the change in ρ is unimportant.

III. RESULTS

The first pair of glasses studied had refractive indices $n_a=1.5056$, $n_b=1.6557$, that is, a borosilicate or phosphate crown with extra dense flint; and the second the indices $n_a=1.5778$, $n_b=1.6267$, that is, medium barium crown and dense flint.

(a) Dispersion Ratio

The changes in ρ necessary to achieve achromatism at various apertures are shown in Figure 1 for these two glass pairs. Although the curvatures of the surfaces and the relative powers of the two components change considerably with aperture, the variation of ρ is very small and is almost the same in both cases.

These curves have been calculated for the standard thickness of one-tenth the aperture for the edge thickness of the crown and the centre thickness of the flint components, and for spherical aberration and coma corrected for the *e* ray ($\lambda=5461 \text{ \AA}$). When other conditions were to be satisfied, the following effects were noted.

(i) *Thickness Change*.—If the standard thickness is increased to 0.15 times the aperture for both components, the change of ρ is negligible, being less than 0.001. However, a slightly larger change is found if the thickness of one component only is altered: the crown thickness remaining constant, an increase of the flint thickness from 0.1 to 0.15 of the aperture decreases ρ by 0.003.

(ii) *Change of Wavelength of Correction*.—If it is desired to correct the spherical aberration and coma for a wavelength *x* different from 5461 \AA , the change required in ρ may be found with sufficient accuracy from the equation

$$\delta\rho \simeq P_{xe} \frac{\rho-1}{\nu_b}, \quad \dots\dots\dots (3)$$

where P_{xe} is the relative partial dispersion between the *e* ray and the new *x* ray,

$$P_{xe} = \frac{n_x - n_e}{n_F - n_C}.$$

It is sufficient to take an average value of P_{xe} for all glasses.

If the aberrations are corrected for the d line,

$$P_{de} \simeq -0.238 \text{ and } \delta\rho = -0.007, \text{ pair I,} \\ = -0.004, \text{ pair II.}$$

When the Brown and Smith tables are used with refractive indices corresponding to a wavelength other than that of the d ray, a similar correction must be applied to ρ , this time taking the partial dispersion from the d line for which these tables are calculated.

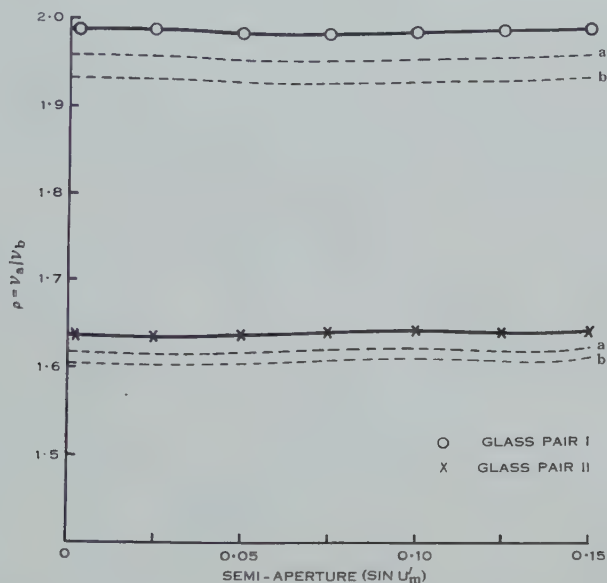


Fig. 1.—The variation of relative dispersion with aperture. Full curves—fully corrected objective; broken curves—objective corrected for use with a prism, (a) of HC glass, (b) of LF glass.

(b) *Maximum Focal Length*

At each aperture, the scale to which the objective may be made and hence the permissible focal length is limited by the tolerances for the residual aberrations. The aberrations that are important and their tolerances are as follows.

(i) *Zonal Spherical Aberration*.—This was found by calculating ΣW_p , the fourth order path difference for the marginal aperture. Its tolerance is 6λ .

(ii) *Zonal Coma*.—This was calculated as the “offence against the sine condition” for the marginal ray. The tolerance for this is $0.53\lambda/h' \sin U'_m$ (Maréchal 1947), the field size h' being chosen so that the astigmatism and field curvature at the edge of the field were equal to one-half their combined tolerance. It was found that zonal coma was negligible in comparison with the other aberrations and it has therefore not been included in Table 1.

(iii) *Secondary Chromatic Aberration*.—This was calculated as the difference between the marginal optical path for e rays and that for C or F rays, these latter rays having the same path length if the primary aberration is corrected. Conrady (1929) gives the tolerance for secondary chromatic aberration as $1\frac{1}{4}$ times the focal range tolerance, that is, $5\lambda/8$ in terms of path difference.

(iv) *Spherochromatic Aberration*.—This was calculated as $\Sigma E_m - \Sigma E_p$ the difference between the marginal and paraxial path differences for C and F rays. Conrady (1923) gives a tolerance of 2λ for this difference.

When each of these aberrations is equal to its tolerance, the maximum focal lengths obtained are given in Table 1, the case of coma being omitted. The value in column (v) is the maximum focal length when all aberrations are taken

TABLE 1

FOCAL LENGTHS AT WHICH SECONDARY ABERRATIONS ARE EQUAL TO TOLERANCE

(i) Zonal spherical aberration, (iii) secondary colour, (iv) spherochromatic aberration, (v) all aberrations considered

Semi-aperture	Maximum Focal Length (mm)							
	Glass Pair I				Glass Pair II			
	(i)	(iii)	(iv)	(v)	(i)	(iii)	(iv)	(v)
$\sin U'_m$								
0.025	10^7	2200	10^5	2200	10^7	2300	10^5	2300
0.050	10^5	540	6000	540	10^5	590	4000	590
0.075	10^4	240	1200	230	10^4	260	910	250
0.100	1500	140	410	130	1500	150	290	130
0.125	420	88	170	77	430	95	120	71
0.150	150	60	86	47	150	65	62	45

into account, all being assumed independent. At the focal length given in column (v) the maximum angular semi-field is limited by astigmatism and curvature, and, in all cases, a semi-field of about 0.012 radian is found to give the mean focus at the edge of the field at a distance of half the focal range tolerance from the best focus on the axis.

It is seen that there are only small differences between the maximum permissible focal lengths for the two glass pairs. The zonal spherical aberrations are practically identical; this agrees with Moffitt and Kaspereit's (1925) findings that, provided the index difference between the crown and the flint is greater than 0.04 , a further increase in index difference has little effect. With regard to secondary colour, the glasses of pair I have typical relative partial dispersions while pair II was chosen to give the maximum practical improvement of this aberration that is obtainable from tabulated glass lists, yet the improvement obtained is very small. The spherochromatic aberration is decreased by an increase of index difference but this aberration is generally less serious than the secondary colour.

(c) *Tolerances for Dispersion*

The effects of variations of ν_a or ν_b and hence of ρ from the ideal value given in Figure 1 can be considered in two ways.

(i) If the doublet is kept exactly aplanatic, the chromatic aberration is given by

$$\left. \begin{aligned} \Sigma E_p &\simeq y^2 A / 2 \\ Lch'_p &\simeq l'^2 A, \end{aligned} \right\} \dots\dots\dots (4)$$

where l' is the image distance and

$$A = \frac{(n_d - 1)_a c_a}{\nu_a} + \frac{(n_d - 1)_b c_b}{\nu_b},$$

c_a and c_b being the total curvatures of the two components. For small changes in ρ from the ideal value,

$$\delta A \simeq - \frac{\delta \rho}{\nu_a f' (\rho - 1)}, \dots\dots\dots (5)$$

and, from this, tolerances for ρ can be calculated from the tolerance for the chromatic aberration.

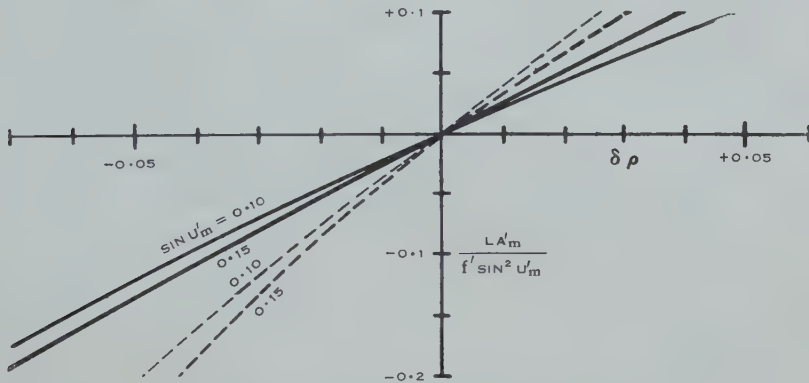


Fig. 2.—The change in spherical aberration due to a change in ρ : Full curves—glass pair I; broken curves—glass pair II.

(ii) Alternatively, the lens can be designed to be chromatically corrected and the spherical correction allowed to change within the tolerance. The variation of spherical aberration with ρ cannot be expressed simply but Figure 2 gives the curves found for this relation during the trigonometrical correction of the objectives, $LA'_m / (f' \sin^2 U'_m)$ being plotted against $\delta \rho$. From these curves the tolerance for ρ can be calculated from the spherical aberration tolerance.

As the latter procedure allows the larger variation of ρ , it is the method that should be used in practice. In Table 2 the permissible variation $\delta \rho$ is given for each glass pair at its maximum focal length (column (v) of Table 1). For other focal lengths $\delta \rho / f'$ can be taken as constant. In this table it is seen that $\delta \rho$ varies approximately as the inverse square of the aperture.

The variation of ρ in Figure 1 from zero aperture to the highest aperture at which a doublet would be used is about 0.005, so it is seen that the same pair of glasses could be used at all apertures.

The strictest tolerance for ρ of 0.01 corresponds to tolerances $\delta v_a = 0.4$ or $\delta v_b = 0.2$ for the dispersions of the two glasses. As most optical glasses are supplied to the tolerance $\delta v = \pm 0.5$, it is only when $\sin U'_m > 0.125$ that a closer specification of dispersion would be necessary.

TABLE 2

TOLERANCES FOR THE RELATIVE DISPERSION $v_a/v_b = \rho$ FOR OBJECTIVES OF MAXIMUM FOCAL LENGTH

Sin U'_m			0.025	0.050	0.075	0.100	0.125	0.150
$\delta\rho$, pair	I	0.5	0.12	0.07	0.04	0.03	0.02
	pair II	0.3	0.07	0.04	0.03	0.02	0.01

IV. PARTIAL CORRECTION

In many applications the doublet objective is not completely corrected but is used to correct the aberrations of other components such as eyepieces or prisms. In this case a different pair of glasses may be required.

(i) *Chromatic Correction*.—If the longitudinal chromatic aberration is not corrected completely, the value of ρ required can be found from equation (4), or by equation (5) for small variations from the corrected value. For under-corrected chromatic aberrations, where Lch'_p or ΣE_p is positive, a decrease of ρ is required.

(ii) *Spherical Correction*.—The change required in ρ for a given spherical aberration is given in Figure 2. It is seen that under-corrected (positive) spherical aberration demands an increase in ρ .

(iii) *Prisms*.—The most common case of incomplete correction is the use of the objective with a prism. This demands an objective under-corrected for both spherical and chromatic aberration; the chromatic effect predominates and a decrease in ρ is required.

In Figure 1 the broken curves represent the values of ρ required for hard crown and light flint prisms of thickness one-fifth the focal length of the objective, using the glasses HC519604 and LF578407. ρ is found to vary linearly with prism thickness.

V. CHOICE OF GLASS

As the variations of ρ over the range of apertures and thicknesses considered are less than the tolerance for ρ , the methods of choosing glass for low aperture objectives referred to in the introduction can be used directly for any aperture. We shall represent the possible glass pairs graphically, using the Brown and Smith (1946) tables.

In Figure 3 the crown region of interest is plotted in terms of n_c and v and superimposed are the curves for glasses that would form doublets with four typical flints, which are taken as Chance LF578407, DF605380, DF623360, and EDF651336. Each curve is shaded to one side to represent the small

increase in ρ with aperture and the curves corresponding to light flint prisms of thickness one-fifth the objective focal length are shown. The curves are terminated at an index difference of 0.04 between crown and flint to avoid high zonal spherical aberrations.

It is seen that there is no suitable glass for use with light flints. Two glasses lie near the dense flint curves, Schott PSK3 and Chance MBC576594, while there is a considerable selection of glasses that can be used with the extra dense flint.

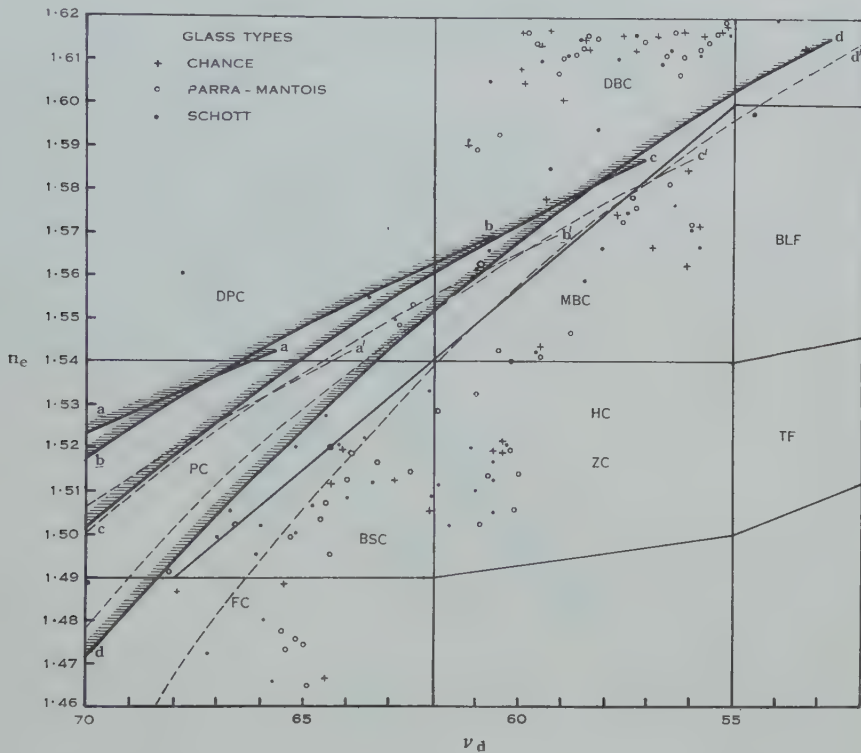


Fig. 3.—Plot of crown glass types suitable for use with the flint glasses: (a) LF578407, (b) DF605380, (c) DF623360, (d) EDF651336. Broken curves show change in glass type when the objective is used with a prism of LF glass.

In Figure 4 the curves for PSK3 and MBC576594 are shown in the region of flint glasses. Both these glasses give close fits to several flints, the former throughout the dense flint range and the latter for the top of the dense flints and the lower extra dense flints.

The curves for the two glasses BSC510644 and MBC572577 which are commonly used for doublets in combination with extra dense flint and dense flint respectively, are shown also in Figure 4. It is seen that these glasses only give perfect correction with a high index extra dense flint, $n \approx 1.69$. But they can be combined with the usual flints when the objective is used with thick prisms.

VI. CONCLUSIONS

The representative cases studied have shown that the glass properties required for cemented doublets which are simultaneously aplanatic and achromatic are the same for lenses of any aperture and that the desirable glasses may be found from tables calculated for thin objectives of small aperture.

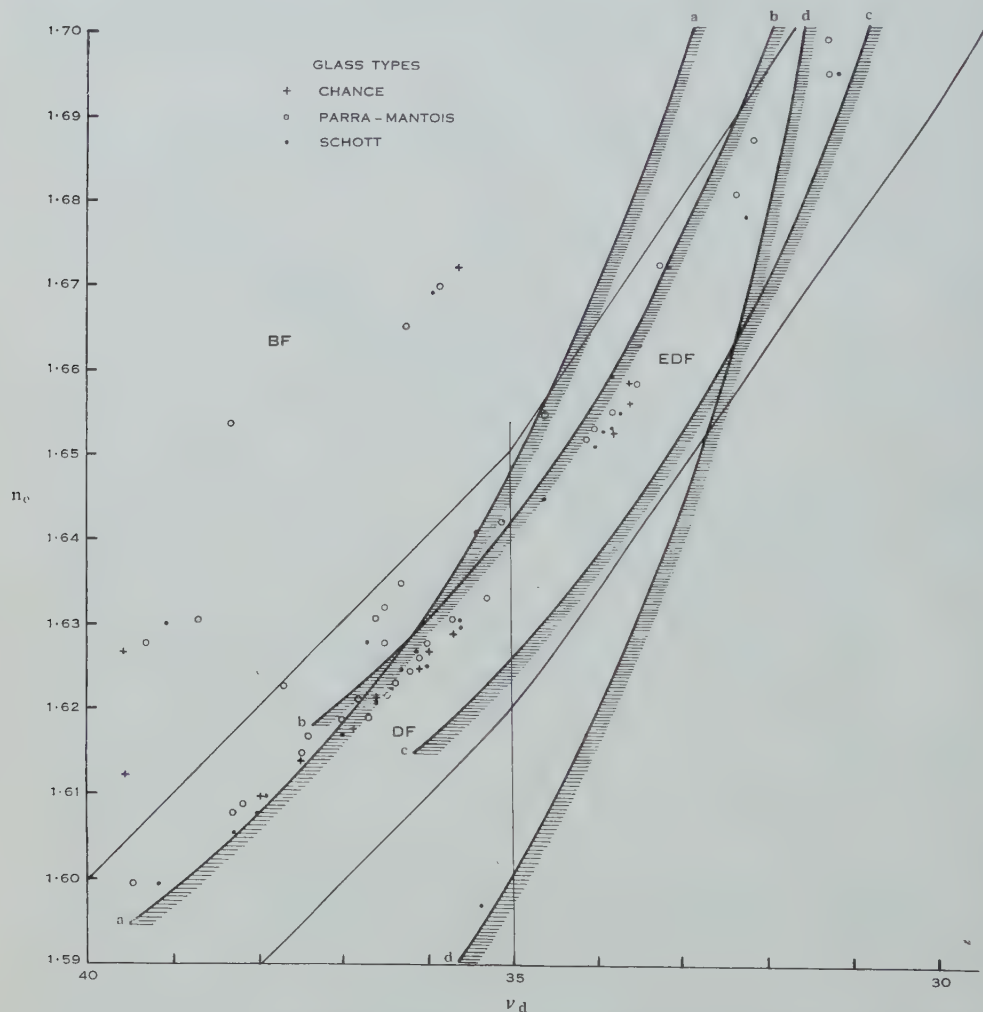


Fig. 4.—Plot of flint glass types suitable for use with the crown glasses : (a) PSK3, (b) MBC576594, (c) MBC572577, (d) BSC510644.

However, the curvatures of the lens surfaces cannot be found from these tables as they vary considerably as the aperture is increased. The greatest change occurs in the curvature of the contact surface which decreases at high apertures.

Tolerances on the glass dispersions have been calculated and at low apertures they are very broad. They are more severe, however, at high apertures and the more commonly used pairs of glasses do not lie sufficiently close to the ideal

values to be suitable. More satisfactory glass types have been indicated; the high order aberrations, however, vary so little between different possible combinations that a final choice would probably depend on questions of stability and transparency of the glasses.

If the objective is used with a prism, or if the "flint in front" form is employed, the difference between the v values of the two components should be reduced, and the common types of crown become the most suitable.

VII. ACKNOWLEDGMENT

The ray traces for this work were performed by Miss J. Ward to whom the author expresses his thanks.

VIII. REFERENCES

- BROWN, E. D., and SMITH, T. (1946).—*Phil. Trans. A* **240**: 59.
CONRADY, A. E. (1904).—*Mon. Not. R. Astr. Soc.* **64**: 182.
CONRADY, A. E. (1923).—In "Dictionary of Applied Physics." (Edited by Sir Richard Glazebrook.) Vol. 4. p. 227. (Macmillan & Co.: London.)
CONRADY, A. E. (1929).—"Applied Optics and Optical Design." (Oxford Univ. Press.)
HARTING, H. (1898).—*Z. InstrumKde.* **18**: 357.
MARÉCHAL, A. (1947).—*Rev. Opt. (Théor. Instrum.)* **26**: 257.
MOFFITT, G. W. (1925).—*J. Opt. Soc. Amer.* **11**: 147.
MOFFITT, G. W., and KASPEREIT, O. K. (1925).—*J. Opt. Soc. Amer.* **11**: 275.

ON THE OPTICAL PROPERTIES OF COMPONENTS FOR BIREFRINGENT FILTERS

By R. G. GIOVANELLI* and J. T. JEFFERIES*

[*Manuscript received December 11, 1953*]

Summary

Expressions are derived for the stray light in a birefringent filter caused by polarizers producing incomplete polarization and by inexact half-wave plates. It is shown that, with suitable orientation of retardation plates, less stray light is obtained with incomplete than with complete polarizers. The asymmetry of the field and the stray light arising from incorrect orientation of the crystal axes are also discussed.

I. INTRODUCTION

Birefringent filters were first developed by Lyot (1933, 1944) and Öhman (1938), the theory being extended later by Billings (1947) and Evans (1949), while Billings, Sage, and Draisin (1951) and Dunn (1951) discussed some of the practical aspects of their construction.

A simple element of such a filter consists of a plane parallel birefringent plate cut with an axis in the surface and mounted between polarizers whose axes bisect those of the crystal. If the plate is of thickness e and birefringence $\varepsilon - \omega$, ε and ω being the refractive indices, the phase retardation κ introduced between the ordinary and extraordinary rays of wavelength λ is given by $\kappa = (\varepsilon - \omega)e/\lambda$, and the transmittance of the element by $\tau = \cos^2 \pi \kappa$ or $\sin^2 \pi \kappa$ respectively for parallel or crossed polarizers. Since κ is a function of λ , the element has a sequence of transmission bands throughout the spectrum.

A simple birefringent filter consists of a series of such elements, each having twice the retardation of the preceding, so that the filter transmits narrow, widely spaced bands (the "principal maxima"). The field, which is limited by the variation in the wavelengths of the principal maxima away from the axis, can be extended in various alternative ways suggested by Lyot. In the one most frequently used (known as Lyot type 1), the retardation plate is split into halves which are separated, with their axes crossed, by an appropriately oriented half-wave plate.

Imperfect components introduce various defects into filter performance. Polarizers which do not produce complete polarization change the fraction of residual light outside the principal maxima, that is, the "stray light". Faulty

* Division of Physics, C.S.I.R.O., University Grounds, Sydney.

half-wave plates in Lyot type 1 elements increase the stray light. Retardation plates of incorrect thickness or imperfect optical quality result in changes of wavelength of the principal maxima. Incorrect orientation of the components with respect to the polarizers increases the stray light, while, if the crystal axes of the retardation plates do not lie in their surfaces, the field is modified. We discuss these matters in some detail.

II. INCOMPLETE POLARIZERS

(a) *General*

Although calcite prism polarizers have high transmission and produce perfect polarization, at least over a field of several degrees, their disadvantages of high cost, bulk, and scarcity have meant that most birefringent filters have been constructed with "Polaroid" polarizers. These latter, however, all transmit a small amount of light vibrating in a direction at right angles to that of the main polarized beam, and so affect the performance of a filter in which they are incorporated. In this section we shall investigate the effects resulting from this incomplete polarization of "Polaroid".

We shall first digress to describe an experiment whose result is of significance in the discussion.

(b) *An Experiment on Retardation in "Polaroid"*

The behaviour of incomplete polarizers in birefringent filters depends in part on any retardation introduced by the polarizer. Let a beam of polarized light be incident on a polarizer, then the beam may be resolved into two components respectively parallel and perpendicular to the polarizer axis and transmitted with unequal absorptions. If there is no phase difference on transmission, the two beams will recombine to give plane polarized light while, if there is a phase difference, elliptically polarized light will result. This effect has been tested on a variety of "Polaroid" films by viewing a sodium lamp through a Thomson prism, a following "Polaroid" sheet being nearly but not quite crossed with it; the yellow sodium glow may then be seen. On rotating a second Thomson prism to examine the light leaving the "Polaroid", extinction occurs at an appropriate orientation, showing that there is no significant phase difference between the transmitted rays. Through some of the older types of "Polaroid" the red filament of the lamp may also be seen, and this disappears at a different orientation of the second Thomson prism, confirming that with such "Polaroids" the ratio of the transmissions respectively parallel and perpendicular to the polarizer axis varies with wavelength.

(c) *Transmission through a Birefringent Filter*

Let unpolarized light be incident on a birefringent element the first polarizer of which has its axis in direction r (Fig. 1) bisecting the axes x and y of the retardation plate. Neglecting losses by reflection at the various surfaces, the vibrations leaving the polarizer may, in virtue of the above, be represented by $r = T_1^{\frac{1}{2}} a \sin \psi$, $s = T_2^{\frac{1}{2}} a \sin \psi$, where ψ is the phase, a is the amplitude of the

incident vibration, and T_1 and T_2 the transmittances of the polarizer for light vibrating respectively along the r and s directions. Considering firstly only the r component incident on the retardation plate, we find that it results in light emerging from the plate with r and s components given by

$$\left. \begin{aligned} r &= T_1^{\frac{1}{2}} a \cos \pi \kappa \sin (\psi - \pi \kappa), \\ s &= \pm T_1^{\frac{1}{2}} a \sin \pi \kappa \cos (\psi - \pi \kappa), \end{aligned} \right\} \dots\dots\dots (2.1)$$

the sign of s depending on the plane of vibration of the ordinary and extraordinary rays in the plate. Similarly, the s component from the first polarizer gives rise to vibrations leaving the retardation plate

$$\left. \begin{aligned} r &= \pm T_2^{\frac{1}{2}} a \sin \pi \kappa \cos (\psi - \pi \kappa), \\ s &= T_2^{\frac{1}{2}} a \cos \pi \kappa \sin (\psi - \pi \kappa). \end{aligned} \right\}$$

On passing through the following polarizer, the r and s amplitudes are multiplied by $T_1^{\frac{1}{2}}$ and $T_2^{\frac{1}{2}}$ respectively if the polarizers are parallel.

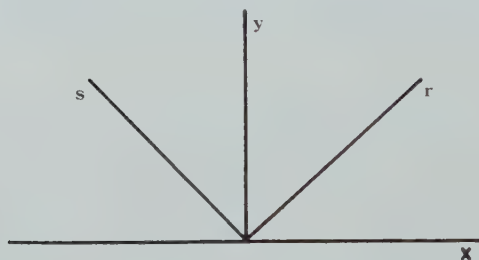


Fig. 1

The beam (2.1) is now incident on a second plate of retardation 2κ , thickness $2e$. The incident r beam behaves as before, its amplitude on striking the retardation plate being $T_1 a \cos \pi \kappa$. The incident s ray, given by (2.1), results in a pair of transmitted components leaving the retardation plate

$$\left. \begin{aligned} r &= \mp T_1^{\frac{1}{2}} T_2^{\frac{1}{2}} a \sin \pi \kappa \sin 2\pi \kappa \sin (\psi - \pi \kappa \mp 2\pi \kappa), \\ s &= T_1^{\frac{1}{2}} T_2^{\frac{1}{2}} a \sin \pi \kappa \cos 2\pi \kappa \cos (\psi - \pi \kappa \mp 2\pi \kappa), \end{aligned} \right\} \dots (2.2)$$

the signs depending on whether the optic axis of the second plate is parallel or at right angles to that of the first plate.

The vibration transmitted through a complete filter is the resultant of the component vibrations which traverse the filter in all possible ways. However, as $T_1 \gg T_2$, terms involving high powers of T_2 may be neglected; these correspond to rays which have passed through several polarizers in the s direction.

Suppose first that all retardation plates are arranged with their optic axes parallel, all polarizers also being parallel. Omitting the phase, the ray passing through all polarizers in the r direction contributes

$$a T_1^{(N+1)/2} \cos \pi \kappa \cos 2\pi \kappa \dots \cos 2^{N-1} \pi \kappa$$

to the emergent vibration, N being the number of elements in the filter. Those making only one transition $r \rightarrow s \rightarrow r$ contribute

$$-aT_1^{N/2}T_2^{\frac{1}{2}} \sum_{l=0}^{N-2} (\cos \pi\kappa \dots \cos 2^{l-1}\pi\kappa \sin 2^l \pi\kappa \sin 2^{l+1}\pi\kappa \dots \cos 2^{l+2}\pi\kappa \dots \cos 2^{N-1}\pi\kappa). \quad (2.3)$$

Those incident and transmitted in the r direction, but traversing two polarizers in the s direction, turn out to be negligible and are omitted. The sum of the above two contributions is denoted as A_1 .

The r contribution due to the incident s component, of random phase with respect to A_1 , is

$$A_2 = aT_1^{N/2}T_2^{\frac{1}{2}} \sin \pi\kappa \cos 2\pi\kappa \dots \cos 2^{N-1}\pi\kappa.$$

The s disturbance due to the ray which has passed through only one polarizer (the last) in the s direction, is

$$A_3 = aT_1^{N/2}T_2^{\frac{1}{2}} \cos \pi\kappa \cos 2\pi\kappa \dots \cos 2^{N-2}\pi\kappa \sin 2^{N-1}\pi\kappa.$$

The transmitted flux is then given by

$$F_\kappa = A_1^2 + A_2^2 + A_3^2. \quad (2.4)$$

(d) The Integrated Flux

For an incident equi-energy spectrum, the transmitted flux is assessed by integrating between two successive principal maxima, with the result

$$F_T = \frac{T_1^{N+1}a^2}{2^N} \left[1 + \frac{T_2}{T_1}(N+1) \right], \quad (2.5)$$

powers of T_2 higher than the first being neglected.

(e) The Flux in a Principal Maximum Band and the Stray Light

With an ideal filter the transmittance falls from unity at the principal maxima, $\kappa = n$, an integer, to zero where $\kappa = n \pm 2^{-N}$, between which lies a principal maximum band. The flux in a principal maximum band, F_M , is found by integrating from $\kappa = n - 2^{-N}$ to $\kappa = n + 2^{-N}$. This integral may be evaluated approximately by putting $\cos x = 1 - x^2/2$, $\sin x = x$, except for the final term of argument $2^{N-1}\pi\kappa$.

Neglecting powers of κ higher than the second, the term in $T_2^{\frac{1}{2}}$, which will be found of special interest, becomes

$$-0.61 \frac{a^2 T_1^{N+\frac{1}{2}} T_2^{\frac{1}{2}}}{2^N}. \quad (2.6)$$

The terms in T_2 integrate to a value of the order of $2^{-N} a^2 T_1^N T_2$, while the term in T_1^{N+1} integrates to $0.89 \times 2^{-N} a^2 T_1^{N+1}$, provided $N > 3$. Hence

$$F_M \simeq \frac{T_1^{N+1}a^2}{2^N} \left[0.89 - 0.61 \left(\frac{T_2}{T_1} \right)^{\frac{1}{2}} + \frac{T_2}{T_1} \right].$$

The fraction of the flux which lies outside the principal maximum band, namely,

$$f_s = \frac{F_T - F_M}{F_T} \simeq 0.11 + 0.61 \left(\frac{T_2}{T_1} \right)^{\frac{1}{2}} + N \frac{T_2}{T_1}, \quad \dots\dots\dots (2.7)$$

we term stray light. With complete polarization, $f_s = 0.11$, a result given previously by Evans (1949).

The lightest available "Polaroid" films have crossed densities of the order of 3, that is, $T_2/T_1 \sim 10^{-3}$. With these, the stray light in a seven-"Polaroid" filter is 0.137. With a crossed density of four, the stray light is reduced to 0.117, but, as the transmitted flux varies as T_1^{N+1} , there is a considerable reduction in transmission if the stray light is to be held to this value.

(f) *The Reduction of Stray Light Using Incomplete Polarization*

With a suitable alignment of retardation plates, an important result can be obtained, namely, a reduction of stray light below that obtained with perfect polarizers.

From (2.2), it can be seen that a ray passing in the $r \rightarrow s \rightarrow r$ direction through two retardation plates with parallel optic axes finishes with a negative amplitude. But if the optic axes of the plates are at right angles (i.e. if the retardation of one is of opposite sign to that of the other), the amplitude is positive. Suppose all polarizers are parallel, but that the optic axis of each plate is perpendicular to that of its immediate predecessor, then A_1 is modified so that the sign of expression (2.6) is reversed, and the stray light is given by

$$f_s \simeq 0.11 - 0.61 \left(\frac{T_2}{T_1} \right)^{\frac{1}{2}} + N \frac{T_2}{T_1}. \quad \dots\dots\dots (2.8)$$

If the element of highest retardation alone is at right angles to the rest, then it can be shown that the stray light is given by

$$f_s \simeq 0.11 - 0.39 \left(\frac{T_2}{T_1} \right)^{\frac{1}{2}} + N \frac{T_2}{T_1}.$$

The optimum reduction in stray light is obtained when

$$\frac{T_2}{T_1} \simeq \frac{0.093}{N^2}, \quad \dots\dots\dots (2.9)$$

when the stray light is reduced to

$$f_s \simeq 0.11 - 0.09/N. \quad \dots\dots\dots (2.10)$$

Thus even with as many as 10 polarizers, the use of the lightest grade of "Polaroid" not only provides for high transmission but also yields a sensible improvement in stray light over that given by calcite polarizers.

(g) The Reduction of Stray Light with Lyot Type 1 Elements

By tracing the changes in the sign of the vibration on passing through Lyot type 1 elements, it is easy to verify that for a reduction in stray light

(1) consecutive Lyot type 1 elements should be arranged with the optic axes of their first components parallel, and

(2) the first component of a Lyot type 1 element should have its optic axis at right angles to that of a preceding or parallel to that of a following simple element.

III. INEXACT HALF-WAVE PLATES

(a) General

For a Lyot type 1 element using parallel polarizers Evans (1949) has given the transmittance in a form reducing to

$$\tau = \cos^2 \pi \kappa_j + \cos^2 \pi \kappa_p (\sin^2 \pi \kappa_j - \sin^2 \pi \delta_j),$$

where κ_p is the retardation of the nominal half-wave plate, and

$$\kappa_j = \kappa_m + \kappa_q,$$

$$\delta_j = \kappa_m - \kappa_q,$$

κ_m and κ_q being the retardations of the two split plates, which we suppose to have equal thicknesses.

As shown by Evans, the effective retardation of a single plate is

$$\kappa = \kappa_0 \left[1 + \frac{\varphi^2}{2\omega} \left(\frac{\cos^2 \theta}{\varepsilon} - \frac{\sin^2 \theta}{\omega} \right) \right], \text{ for } \varepsilon > \omega, \dots\dots (3.1)$$

κ_0 being the retardation for axial passage and φ and θ respectively the angles of incidence and the azimuth of the plane of incidence measured from the fast axis. Since the axes of the two plates are crossed, a ray at φ, θ to the first slab will enter the second at $\varphi, \theta + \pi/2$.

Then

$$\kappa_j = \kappa_{j0} \left[1 + \frac{\varphi^2}{4\omega} \left(\frac{1}{\varepsilon} - \frac{1}{\omega} \right) \right], \dots\dots\dots (3.2)$$

and

$$\delta_j = \kappa_{j0} \left[\frac{\varphi^2}{4\omega} \left(\frac{1}{\varepsilon} + \frac{1}{\omega} \right) \cos 2\theta \right],$$

$$\kappa_p = \kappa_{p0} \left[1 + \frac{\varphi^2}{4\omega} \left\{ \left(\frac{1}{\varepsilon} - \frac{1}{\omega} \right) - \left(\frac{1}{\varepsilon} + \frac{1}{\omega} \right) \sin 2\theta \right\} \right].$$

The variation in κ_j with φ normally sets a limit to the field, involving a variation in λ_M , the wavelength of the principal maximum. If $\Delta\lambda_M$ is not to exceed 0.1 of the half-width of the principal band,

$$\Delta\kappa_j \leq 2^{-N}/10,$$

which when combined with (3.1) yields the maximum value of φ .

An inexact half-wave plate introduces stray light, which we now discuss.

(b) *The Integrated Flux*

The flux transmitted by a birefringent filter of N elements, the first J of which are unsplit and the remainder Lyot type 1 elements, is

$$F = a^2 \int \cos^2 \pi \kappa \cos^2 2\pi \kappa \dots \cos^2 2^{J-1} \pi \kappa G(2^J \pi \kappa) G(2^{J+1} \pi \kappa) \dots G(2^{N-1} \pi \kappa) d\kappa,$$

where

$$G(2^x \pi \kappa) = \cos^2 2^x \pi \kappa + \gamma (\sin^2 2^x \pi \kappa - \sin^2 2^x \beta \pi \kappa),$$

$$\gamma = \cos^2 \pi \kappa_p,$$

and, to a good approximation,

$$\beta = \frac{\varphi^2}{4\omega} \left(\frac{1}{\varepsilon} + \frac{1}{\omega} \right) \cos 2\theta.$$

Provided γ is small and varies negligibly over the range of integration, we find, as in Section II, that the flux between two successive principal maxima, F_T , may be written, for $2^{N-1}\beta < 1/4$,

$$F_T = \frac{a^2}{2^N} \left[1 + \gamma \sum_{x=J+1}^N \frac{\cos 2^x \beta \pi \kappa_0 \sin 2^{x-1} \beta \pi}{2^{x-1} \beta \pi} \right].$$

Similarly, with the same approximations, we find the flux in a principal maximum band to be given by

$$F_M = \frac{a^2}{2^N} \left[0.89 + \gamma \left(\frac{2}{3} - \frac{4^{J-N+1}}{9} \right) - \gamma \sum_{x=J+1}^N y_x \sin^2 2^{x-1} \beta \pi \kappa_0 \right],$$

where

$$y_x = \frac{8}{9} + \frac{4^{x-N}}{3}, \quad x \neq N,$$

$$y_N = 1.45.$$

The stray light fraction on the axis is thus

$$f_s = \frac{0.11 + \gamma \left(M - \frac{2}{3} + \frac{4^{1-M}}{9} \right)}{1 + M\gamma},$$

where $M = N - J$ is the number of wide-field elements.

As an example consider a six-element filter designed to operate at the coronal red and green lines. If provided with mica half-wave plates adjusted for the sodium D line, then at the coronal line wavelengths $\gamma \sim 0.03$. The stray light fraction is given in Table 1.

To investigate off-axis effects, suppose the last two elements split. At angles such that $\sin 2^{x-1} \beta \pi \simeq 2^{x-1} \beta \pi$, the stray light is

$$f_s = \frac{0.11 + \gamma (0.57 + 0.51 \cos 2^5 \beta \pi \kappa_0 + 0.28 \cos 2^6 \beta \pi \kappa_0)}{1 + \gamma (\cos 2^5 \beta \pi \kappa_0 + \cos 2^6 \beta \pi \kappa_0)}.$$

This varies around the field through its dependence on β and hence on θ , but *diminishes* away from the axis and so imposes no limit to the field.

IV. INCORRECT ORIENTATION OF COMPONENTS ABOUT A NORMAL TO THE SURFACE

In an otherwise perfect simple element, let the angle between the first polarizer axis r and the fast axis x of the crystal plate be ξ , the final polarizer being at an angle ζ to x . The first polarizer transmits a vibration $r = a \sin \psi$, the vibration transmitted through the whole element being plane polarized and represented by

$$a \cos \xi \cos \zeta \cdot \sin \psi + a \sin \xi \sin \zeta \cdot \sin (\psi - 2\pi\kappa).$$

The transmittance of the element is thus

$$\tau = A \cos^2 \pi\kappa + B \sin^2 \pi\kappa,$$

where $A = \cos^2 (\xi - \zeta)$ and $B = \cos^2 (\xi + \zeta)$.

TABLE 1
STRAY LIGHT AT $\lambda 5303$ WITH MICA HALF-WAVE
PLATES ADJUSTED AT $\lambda 5890$

Number of Split Elements	Stray Light Fraction
0	0.110
1	0.119
2	0.142
3	0.165
4	0.188
5	0.208
6	0.229

The transmittance of a simple birefringent filter is the product of a set of such terms

$$\tau = (A_1 \cos^2 \pi\kappa + B_1 \sin^2 \pi\kappa)(A_2 \cos^2 2\pi\kappa + B_2 \sin^2 2\pi\kappa) \dots \\ (A_N \cos^2 2^{N-1} \pi\kappa + B_N \sin^2 2^{N-1} \pi\kappa).$$

In practice, the A 's will be almost unity and the B 's small, so that terms involving higher powers than B may be neglected. As before, the flux in the whole transmission band is obtained by integrating from $\kappa = n$ to $n+1$, yielding

$$F_T = \frac{\Pi_x}{2^N} \left(1 + \sum_{x=1}^N \frac{B_x}{A_x} \right),$$

where $\Pi_x = A_1 A_2 \dots A_N$.

The flux in a principal maximum band is given by

$$F_M = \frac{\Pi_x}{2^N} \left[0.89 + \frac{1}{3} \sum_{x=1}^{N-1} \frac{B_x}{A_x} 4^{x-N} + \frac{5}{9} \frac{B_N}{A_N} \right].$$

If the ratio B/A is constant and equal to ρ ,

$$F_T = \frac{\Pi_x}{2^N} (1 + N\rho); \quad F_M = \frac{\Pi_x}{2^N} \left(0.89 + \frac{2}{3}\rho \right) \quad \text{for } N > 3.$$

These results show that if $\xi + \zeta = \pi/2$, that is, $B=0$, the effect is simply to reduce the flux uniformly over the whole spectrum ; there is no change in the proportion of scattered light.

With B/A constant, the stray light fraction is

$$f_s = \frac{0.11 + \left(N - \frac{2}{3}\right)\rho}{1 + N\rho},$$

which, for readily achievable orientations and normal numbers of elements, approximates to

$$f_s = 0.11 + N\rho \simeq 0.11 + N\left(\frac{\pi}{2} - \xi - \zeta\right)^2.$$

If the errors in orientation ξ and ζ are all equal to x (radians) and are of the same sign,

$$f_s = 0.11 + 4Nx^2. \quad \dots\dots\dots (4.1)$$

V. MISALIGNMENT OF THE AXES WITH RESPECT TO THE SURFACE OF A PLATE

(a) General Equations

Here the field size of a simple element is unchanged, as shall be shown, but may be displaced with respect to the normal to the surface of the element.

Following Evans's (1949) method, consider a plane parallel uniaxial crystal of thickness e , whose axes form a coordinate system, the x axis of which is parallel to the crystal optic axis. Choose units of time and distance which make the velocity of light *in vacuo* unity. Suppose the crystal surfaces do not lie in a plane containing two of the axes, the equations of the surfaces being

$$\alpha x + \beta y + \gamma z = 0, \quad \dots\dots\dots (5.1)$$

$$\alpha x + \beta y + \gamma z - e = 0, \quad \dots\dots\dots (5.2)$$

where α , β , and γ are the direction cosines of the normals to the surfaces. An incident ray has a wave front

$$ax + by + cz - t = 0, \quad \dots\dots\dots (5.3)$$

where a , b , and c are the direction cosines of the ray.

At the origin, on the first surface, a secondary wavelet spreads into the crystal, its equation being that of the ellipsoid

$$\zeta^2 x^2 + \eta^2 y^2 + \nu^2 z^2 - t^2 = 0, \quad \dots\dots\dots (5.4)$$

where ζ , η , and ν are the reciprocals of the velocities in the x , y , and z directions, that is, the refractive indices.

A tangent plane to the ellipsoid at x_1 , y_1 , z_1 is

$$x_1 \zeta^2 x + y_1 \eta^2 y + z_1 \nu^2 z - t = 0. \quad \dots\dots\dots (5.5)$$

The condition for this to be a wave front is that the three planes (5.1), (5.3), and (5.5) meet along a common line, that is,

$$\begin{vmatrix} \alpha & \gamma & 0 \\ a & c & -t \\ x_1 \zeta^2 & z_1 v^2 & -t^2 \end{vmatrix} = 0 = \begin{vmatrix} \beta & \gamma & 0 \\ b & c & -t \\ y_1 \eta^2 & z_1 v^2 & -t^2 \end{vmatrix},$$

from which

$$\left. \begin{aligned} x_1 &= (\gamma \zeta^2)^{-1} (\alpha v^2 z_1 - \alpha c t + \gamma a t), \\ y_1 &= (\gamma \eta^2)^{-1} (\beta v^2 z_1 - \beta c t + \gamma b t). \end{aligned} \right\} \dots\dots\dots (5.6)$$

Substituting these in (5.4) we find

$$v z_1^2 = t^2 - (\gamma^2 \zeta^2)^{-1} (\alpha v^2 z_1 - \alpha c t + \gamma a t)^2 - (\gamma^2 \eta^2)^{-1} (\beta v^2 z_1 - \beta c t + \gamma b t)^2, \quad (5.7)$$

whose solution is $z_1 = t.F$, where, restricting consideration to small errors in alignment, so that α and β are small and $\gamma \simeq 1$,

$$F = \frac{1}{v} \left(1 - \frac{a^2}{\zeta^2} - \frac{b^2}{\eta^2} + \frac{2a\alpha c}{\zeta^2} + \frac{2b\beta c}{\eta^2} \right)^{\frac{1}{2}} - \frac{\alpha a}{\zeta^2} - \frac{\beta b}{\eta^2}.$$

The time t_1 for the ray to reach the second surface is found by putting $x = x_1$, $y = y_1$, $z = z_1$ in (5.2), so

$$\alpha (\gamma \zeta^2)^{-1} (\alpha v^2 t_1 F - \alpha c t_1 + \gamma a t_1) + \beta (\gamma \eta^2)^{-1} (\beta v^2 t_1 F - \beta c t_1 + \gamma b t_1) + \gamma t_1 F = e,$$

which yields to the first order,

$$\left. \begin{aligned} t_1 &= e \left(F + \frac{\alpha a}{\zeta^2} + \frac{\beta b}{\eta^2} \right)^{-1} \\ &= e v \left(1 - \frac{a^2}{\zeta^2} - \frac{b^2}{\eta^2} + \frac{2a\alpha c}{\zeta^2} + \frac{2b\beta c}{\eta^2} \right)^{\frac{1}{2}}. \end{aligned} \right\} \dots\dots\dots (5.8)$$

The emerging wave front, parallel to (5.3), is

$$ax + by + cz - (t - \Delta) = 0,$$

where

$$\Delta = t_1 - ax_1 - by_1 - cz_1.$$

The distance of the wave from the origin is

$$p = t - \Delta = t - t_1 + ax_1 + by_1 + cz_1. \quad \dots\dots\dots (5.9)$$

For the extraordinary wave,

$$\zeta = \omega, \quad \eta = v = \varepsilon, \quad t_1 = t_\varepsilon.$$

For the ordinary wave,

$$\zeta = \eta = v = \omega, \quad t_1 = t_\omega.$$

Hence we find for the retardation, $\kappa = (p_\omega - p_\varepsilon)/\lambda$,

$$\kappa = \frac{1}{\lambda} \left\{ (F_\omega t_\omega - F_\varepsilon t_\varepsilon) \left(\frac{\beta b}{\gamma} + c \right) + \frac{F_\omega t_\omega \alpha a}{\gamma} - \frac{F_\varepsilon t_\varepsilon \alpha a \varepsilon^2}{\gamma \omega^2} + (t_\omega - t_\varepsilon) \left(\frac{a^2}{\omega^2} - \frac{a \alpha c}{\gamma \omega^2} - 1 \right) + \left(\frac{t_\omega}{\omega^2} - \frac{t_\varepsilon}{\varepsilon^2} \right) \left(b^2 - \frac{b \beta c}{\gamma} \right) \right\},$$

which after some reduction becomes, to the first order,

$$\kappa = \kappa_0 \left\{ 1 - \frac{a^2}{2\omega^2} + \frac{b^2}{2\omega\varepsilon} + \frac{\alpha a}{\omega^2} (c - \omega - \varepsilon) - \frac{b\beta c}{\omega\varepsilon} \right\}. \quad \dots\dots\dots (5.10)$$

(b) The Simple Element

The limit of the field is reached when $\kappa - \kappa_0$ reaches the tolerable limit. In the perfect element, where $\alpha = \beta = 0$,

$$\frac{\kappa - \kappa_0}{\kappa_0} = \frac{b^2}{2\omega\varepsilon} - \frac{a^2}{2\omega^2},$$

which has maxima at $b=0$ and $a=0$, if $b^2 + a^2 = \text{constant}$, and is zero when $b^2/a^2 = \varepsilon/\omega$.

(i) Case 1.—If $a=0$, (5.10) becomes

$$\frac{\kappa - \kappa_0}{\kappa_0} = \frac{b^2}{2\omega\varepsilon} - \frac{\beta b(1-b^2)^{\frac{1}{2}}}{\omega\varepsilon} \quad (\text{as } a^2 + b^2 + c^2 = 1).$$

For a given change in retardation, let $b=b_0$ for $\beta=0$, and $b=B$ for $\beta \neq 0$. Then

$$b_0^2 = B^2 - 2\beta B(1-B^2)^{\frac{1}{2}},$$

which, for $\beta \ll b_0$, and B small, reduces to

$$B - \beta = \pm b_0.$$

As the angle which the limiting ray makes with the plane containing the optic axis and the normal to the surface is given closely by $B - \beta$, the field in the plane $x=0$ is unchanged and symmetrical.

(ii) Case 2.—If $b=0$, the corresponding result is

$$A - \alpha = \pm a_0 - \alpha(\omega + \varepsilon).$$

The angle which the limiting ray makes with the plane containing the y -axis and the normal to the surface is given closely by $A - \alpha$, so that the field in the plane $y=0$ is unchanged in size but asymmetrical by an amount $\alpha(\omega + \varepsilon)$.

(c) Lyot Type 1 Element

The large field of a perfect Lyot type 1 element results from a compensation of off-axis effects. We shall now show that this compensation is reduced and the field size diminished if the axes of the two plates are not parallel to the surfaces.

Suppose that axes of the first plate of a Lyot type 1 element are not parallel to the surface of the plate, the second plate being free from error.

With respect to the coordinate system established by the crystal axes of the first plate, those of the second plate have direction cosines $(\alpha_1, \beta_1, \gamma_1)$, $(\alpha_2, \beta_2, \gamma_2)$, and $(\alpha_3, \beta_3, \gamma_3)$, where $(\alpha_3, \beta_3, \gamma_3) = (\alpha, \beta, \gamma)$ are the direction cosines of the normal to the surface. Hence the direction cosines of a ray which in the first system are (a, b, c) become in the coordinate system of the second plate (a', b', c') where

$$\begin{aligned}a' &= \alpha_1 a + \beta_1 b + \gamma_1 c, \\b' &= \alpha_2 a + \beta_2 b + \gamma_2 c, \\c' &= \alpha_3 a + \beta_3 b + \gamma_3 c.\end{aligned}$$

The retardation of the complete element is thus

$$\kappa = \kappa_1 \left\{ 1 - \frac{a^2}{2\omega^2} + \frac{b^2}{2\omega\varepsilon} + \frac{\alpha a}{\omega^2}(c - \omega - \varepsilon) - \frac{b\beta c}{\omega\varepsilon} \right\} + \kappa_2 \left\{ 1 + \frac{a'^2}{2\omega\varepsilon} - \frac{b'^2}{2\omega^2} \right\}.$$

Since $\alpha_1 \simeq 1$, β_1 and γ_1 are very small, $\beta_2 \simeq 1$, α_2 and γ_2 are very small, and $c \simeq 1$, it readily follows that $\gamma_1 \simeq -\alpha$ and

$$\begin{aligned}\kappa &= \kappa_1 \left\{ 1 - \frac{a^2}{2\omega^2} + \frac{b^2}{2\omega\varepsilon} + \frac{\alpha a}{\omega^2}(1 - \omega - \varepsilon) - \frac{\beta b}{\omega\varepsilon} \right\} \\&\quad + \kappa_2 \left\{ 1 + \frac{a^2 + 2a(\beta_1 b - \alpha)}{2\omega\varepsilon} - \frac{b^2 - 2b(\beta_1 a + \beta)}{2\omega^2} \right\}.\end{aligned}$$

Suppose the plates are adjusted during construction so that in each the retardation along the normal is equal to $\frac{1}{2}K$. In the first plate, the direction of the normal is $a = \alpha$, $b = \beta$; in the second, $a' = 0 = b'$. Thus very closely

$$\begin{aligned}\kappa_1 &= \frac{1}{2}K \left\{ 1 - \frac{\alpha^2}{\omega^2}(\frac{1}{2} - \omega - \varepsilon) + \frac{\beta^2}{2\omega\varepsilon} \right\}, \\ \kappa_2 &= \frac{1}{2}K,\end{aligned}$$

whence very closely

$$\begin{aligned}\frac{\kappa - K}{K} &= \frac{1}{2} \left\{ -\frac{\alpha^2}{\omega^2}(\frac{1}{2} - \omega - \varepsilon) + \frac{\beta^2}{2\omega\varepsilon} - \frac{a^2}{2\omega^2} + \frac{b^2}{2\omega\varepsilon} + \frac{\alpha a}{\omega^2}(1 - \omega - \varepsilon) - \frac{\beta b}{\omega\varepsilon} + \frac{a^2 + 2a(\beta_1 b - \alpha)}{2\omega\varepsilon} \right. \\&\quad \left. - \frac{b^2 - 2b(\beta_1 a + \beta)}{2\omega^2} \right\}. \dots\dots\dots (5.11)\end{aligned}$$

(i) *Case 3.*—If $a = 0$, (5.11) becomes

$$\frac{\kappa - K}{K} = \frac{1}{2} \left\{ -\frac{\alpha^2}{\omega^2}(\frac{1}{2} - \omega - \varepsilon) + \frac{\beta^2}{2\omega\varepsilon} + \frac{1}{\omega^2\varepsilon}(\omega - \varepsilon) \left(\frac{b^2}{2} - \beta b \right) \right\}.$$

For a given change in retardation, let $b = b_0$ for $\alpha = \beta = 0$, and $b = B$ for $\alpha \neq 0$, $\beta \neq 0$. Then it readily follows that

$$B - \beta = \pm \left\{ b_0^2 + \beta^2 - \frac{\beta^2 \omega - 2\alpha^2 \varepsilon (\frac{1}{2} - \omega - \varepsilon)}{\omega - \varepsilon} \right\}^{\frac{1}{2}}. \dots\dots (5.12)$$

Hence the field is symmetrical and reduced.

(ii) *Case 4.*—If $b=0$, the corresponding result is

$$A - \alpha = \frac{\alpha \varepsilon (\omega + \varepsilon)}{\omega - \varepsilon} \pm \left\{ a_0^2 + \alpha^2 \left[\frac{\varepsilon (1 - \omega - \varepsilon) - \omega}{\omega - \varepsilon} \right]^2 - \frac{\beta^2 \omega - 2\alpha^2 \varepsilon (\frac{1}{2} - \omega - \varepsilon)}{\omega - \varepsilon} \right\}^{\frac{1}{2}} \\ \simeq \frac{4 \cdot 5 \alpha}{\omega - \varepsilon} \pm \left\{ a_0^2 + \left(\frac{4 \cdot 5 \alpha}{\omega - \varepsilon} \right)^2 - \frac{1 \cdot 5 \beta^2 + 7 \cdot 5 \alpha^2}{\omega - \varepsilon} \right\}^{\frac{1}{2}} \dots \dots \dots (5.13)$$

Hence in this plane the field is unsymmetrical by about $4 \cdot 5 \alpha / (\omega - \varepsilon)$, and since $(\omega - \varepsilon)^{-1}$ is -110 for quartz and $5 \cdot 8$ for calcite, the asymmetry will be large unless α is very small indeed, that is, unless the optic axis lies almost exactly in the surface.

Unless $\beta \gg \alpha$, the field is increased from $2a_0$ to approximately $2[a_0^2 + \{4 \cdot 5 \alpha / (\omega - \varepsilon)\}^2]^{\frac{1}{2}}$, the minimum angle between the edge of the field and the normal being $[a_0^2 + \{4 \cdot 5 \alpha / (\omega - \varepsilon)\}^2]^{\frac{1}{2}} - 4 \cdot 5 \alpha / (\omega - \varepsilon)$. This is $0 \cdot 4 a_0$ when $\alpha = a_0(\omega - \varepsilon) / 4 \cdot 5$, or $\alpha \simeq a_0 / 500$ for quartz and $a_0 / 25$ for calcite.

Nor may β be allowed to be very large. From both (5.12) and (5.13) it is evident that if $\alpha = 0$, there is a 10 per cent. reduction in field if $\beta \simeq (\omega - \varepsilon / 5 \omega)^{\frac{1}{2}} b_0$, or about $b_0 / 30$ for quartz or $b_0 / 7$ for calcite.

It should be noted that the field is given by the above expressions only when $4 \cdot 5 \alpha / (\omega - \varepsilon)$ is not too large; for large enough values, the retardation in the centre of the asymmetric field, which is opposite in sign to that at the edge, also exceeds the tolerable limits.

VI. DISCUSSION

(a) *Stray Light*

(i) *Defective Polarizers.*—In the design of a birefringent filter, the question arises as to the type of polarizer, calcite or "Polaroid", and if the latter, what shall be its characteristics. While calcite prisms undoubtedly give good results, properly chosen "Polaroid" will often be preferable because of availability and cheapness. Usually, high transmission is an important factor; but this is obtainable in "Polaroid" only with incomplete polarization. However, by alternating the orientation of the retardation plates, the stray light can be reduced. A good compromise in a seven-"Polaroid" filter would be obtained with "Polaroid" for which $T_1/T_2 = 500$. Two pieces of this when parallel have a transmittance 250 times that when crossed. Such material would normally have a parallel transmittance (almost exactly $\frac{1}{2} k^2 T_1^2$) of about $0 \cdot 40$ —the term k allowing for surface reflections—and a seven-"Polaroid" filter would have a peak transmittance 82 per cent. of that of a similar filter equipped with calcite polarizers. The stray light would be reduced from $0 \cdot 11$ to $0 \cdot 10$. These results are, however, simplifications derived on the assumption of constant T_2/T_1 , which in practice varies throughout the spectrum.

(ii) *Half-wave Plates.*—A second design consideration concerning the introduction of stray light is the use of non-achromatic half-wave plates in Lyot type 1 wide-field elements. If the filter is intended for a range of wavelengths,

and mica half-wave plates are adjusted for the middle of the range, the stray light introduced depends on the range of wavelengths and number of wide-field elements, and can be serious. Even in a filter designed for a single wavelength and containing several wide-field elements the tolerance on the accuracy of the half-wave plate is quite exacting.

(iii) *Alignment between Retardation Plates and Polarizers.*—A further source of stray light is purely constructional, namely, the incorrect alignment between the retardation plates and polarizers. Its influence will be negligible if the errors are of the order of 0.5° or less.

We have not discussed the effects of errors in thickness of retardation plates; these have been treated by others, a tolerance of $\lambda/20$ in retardation being normally acceptable.

(b) *Field*

The major effect of faulty components on the field is an asymmetry due to lack of alignment of the optic axis in the surface of an element. The effect is particularly marked in the case of Lyot type 1 elements, where the asymmetry is some 500 times the error in alignment for quartz and 17 times for calcite.

Errors in the other crystal axes are not very important in simple elements; in Lyot type 1 elements, the field is reduced by some 10 per cent. for an alignment error of about 3 per cent. of the semi-field for quartz or 15 per cent. of the semi-field for calcite.

In Lyot type 1 elements, one way of compensating for loss of field is to provide in the filter cell adjustment for aligning the individual components. Since the field of a simple element is asymmetrical by about 3α , the available adjustment should be at least three times the maximum error expected in aligning the crystal axes in the plane of the surface. However, the large asymmetry in field introduced by errors in alignment provides a means of checking this alignment during construction of the elements, and in this way avoiding the necessity for added complexity in the filter cell.

VII. REFERENCES

- BILLINGS, B. H. (1947).—*J. Opt. Soc. Amer.* **37**: 738.
BILLINGS, B. H., SAGE, S., and DRAISIN, W. (1951).—*Rev. Sci. Instrum.* **22**: 1009.
DUNN, R. B. (1951).—*Sky & Telesc.* **10**: 148.
EVANS, J. W. (1949).—*J. Opt. Soc. Amer.* **39**: 229.
LYOT, B. (1933).—*C.R. Acad. Sci., Paris* **197**: 1593.
LYOT, B. (1944).—*Ann. Astrophys.* **7**: 31.
ÖHMAN, Y. (1938).—*Nature* **141**: 291.

DIPOLE RESONANT MODES OF AN IONIZED GAS COLUMN

By R. E. B. MAKINSON* and D. M. SLADE*

[Manuscript received December 3, 1953]

Summary

It is shown by a quasi-static treatment there is not one but a multiplicity of plasma resonant modes which will emit dipole radiation in a cylindrical column of ionized gas with a radially symmetrical but non-uniform electron density. It is shown that the multiple resonances obtained by Romell in scattering of short waves from a gas discharge column may be qualitatively explained assuming a reasonable electron distribution. Approximate values for the resonant frequencies for a Gaussian distribution, relevant to meteor trails, are found and some of these may be observable.

The status of the finite energy loss and phase shift, shown by Herlofson and other authors to arise where the real part of the dielectric constant vanishes, is examined and mechanisms pointed out which modify the expressions given.

I. INTRODUCTION

It was shown by Tonks (1931) and emphasized by Herlofson (1951) that the frequencies of plasma oscillations depend on the shape of the boundaries and that putting a dielectric constant

$$K \equiv 1 - ne^2/(\pi m \nu^2) = 0,$$

where n is an electron density, does not always serve to locate these frequencies. Not much study has yet been given to the oscillations occurring with non-uniform densities; it does not, for example, appear to be realized that in this case there will in general be a number of dipole modes of different frequencies apart from a multiplicity of quadrupole and higher modes. This is shown in the following paragraphs and related to the experiments of Romell (1951) where such modes were found in reflection from a gas discharge column.

Closer examination is also needed of the paradoxical conclusion that with vanishing collision frequency there is still a finite energy loss and phase shift (Makinson *et al.* 1951; Kaiser and Closs 1952) arising near the region where the real part of K goes through zero.

II. THE FIELD EQUATIONS

If the quantities varying with time have frequency ν , the electron density is $n_0(\mathbf{r}) + n(\mathbf{r}, t)$ and the force on an electron is taken as $-Ee/m$ to the first order,

* School of Physics, University of Sydney.

we obtain from Maxwell's equations when streaming and thermal velocities are supposed zero :

$$\operatorname{div} \mathbf{E} = -4\pi ne, \quad \dots \quad (1)$$

$$\operatorname{div} (K\mathbf{E}) = 0, \quad K = 1 - \frac{n_0 e^2}{\pi m \nu^2}, \quad \dots \quad (2)$$

$$K \operatorname{div} \mathbf{E} = \frac{e^2}{\pi m \nu^2} \mathbf{E} \cdot \operatorname{grad} n_0, \quad \dots \quad (3)$$

$$\frac{\partial^2 n}{\partial t^2} = -\frac{4\pi n_0 e^2}{m} n + \frac{e}{m} \mathbf{E} \cdot \operatorname{grad} n_0, \quad \dots \quad (4)$$

so that K is the dielectric constant of the plasma. We will neglect at first energy losses due to collisions; this is known to be a useful approximation in many problems. The positive ions in the plasma are effectively stationary at the frequencies of interest and the mean net charge density is taken as zero.

(a) Uniform Density

It follows from (2) and (3) that if n_0 is uniform over a region, either (i) $\operatorname{div} \mathbf{E} = 0$ over that region so that, in any forced oscillations produced by external fields of arbitrary frequency, the electron cloud moves without density changes, or (ii) $K = 0$ and oscillations may take place only at the corresponding (Langmuir) frequency $\nu_L = (n_0 e^2 / \pi m)^{1/2}$, there being oscillation of the electron density. The name "plasma electron resonance" was given by Tonks (1931) to such motion. Both types may coexist at the Langmuir frequency.

In case (i) there are restoring forces produced at the boundaries of the plasma, and there are resonant frequencies in general quite different from ν_L which depend on the shape of the boundaries and the orientation and order of the multipole electric moment produced. Tonks called these plasma resonances. They are discussed also by Herlofson (1951).

If the free-space wavelength is much greater than the extent of the plasma, as will now be assumed, we may find such plasma-resonant frequencies simply from the well-known expressions in electrostatics by noting the values of K for which the field inside the region is infinite, that is, finite for vanishing external field (Kaiser and Closs 1952). If the external field is uniform, frequencies of the dipole modes result; from the expressions for a non-uniform field, frequencies of the multipole modes as well.

(b) Non-uniform Density

Except in the special case where \mathbf{E} is everywhere perpendicular to $\operatorname{grad} n_0$, $\operatorname{div} \mathbf{E} \neq 0$, the density no longer remains constant and we lose any such distinction as between types of oscillation (i) and (ii) above. The Langmuir frequency varies from point to point and has no special meaning.

The quasi-static treatment is again applicable to find resonant frequencies but there is no body of standard electrostatic results for non-uniform media to draw upon.

Since $\text{div } \mathbf{D} = 0$, $\mathbf{D} = K\mathbf{E}$, continuous lines of electric displacement may be drawn through such a plasma, their spacing giving the magnitude of \mathbf{D} . It is now necessary to suppose the fields sinoidal and to give K the complex form $K_1 - jK_2$, the imaginary term arising from energy losses due to processes such as collisions, cf. Section V. We suppose K_2 very small, consequently unimportant except near where $K_1 = 0$. Near a surface where K_1 goes through zero, E changes suddenly from large positive to large negative values and a phase shift is introduced, while n becomes very large according to equations (1)–(4). There is an oscillatory piling up of charge towards and away from this surface, with smaller piling up throughout the plasma. The restoring forces governing the frequency of a particular resonant mode arise from the free charges appearing at the boundaries as well as from such piling up.

III. DIPOLE MODES OF NON-UNIFORM CIRCULAR CYLINDER

If the cylinder has axial symmetry and the vanishingly small applied uniform electric field is transverse, putting $\mathbf{E} = -\text{grad } V$, $V = R(r) \cos \theta$ in cylindrical coordinates, R must satisfy (Kaiser and Closs 1952)

$$R'' + \left(\frac{K'}{K} + \frac{1}{r} \right) R' - \frac{R}{r^2} = 0, \quad \dots \dots \dots (5)$$

with the boundary condition for resonance (in the case where $K \rightarrow 1$ as $r \rightarrow \infty$) $R(0) = 0$, real part of $R \sim \text{const.}/r$ for large R . If we put $K = 1 - \lambda f(r) - jK_2$ so that $f(\infty) = 0$, only for certain eigenvalues of λ can the boundary conditions be satisfied. We now show that there can be no eigenvalues (and hence no plasma resonance) unless K is negative for some range of R .

From (5) and the condition $R(0) = 0$, if R is, say, positive at $r = 0$, R can never turn downward if K is everywhere positive, since at such turning point we would require R'' negative which is contrary to (5). Thus R increases monotonically and is proportional to r , not r^{-1} , for large r . If, however, K_1 passes through zero, R turns down again at that point in such a way that it *may* for some values of λ behave at infinity like r^{-1} .

It is apparent from (5) that if $K(r)$ is replaced by $K(mr)$, $R(r)$ by $R(mr)$, where m is any multiplier, the eigenvalues will be unchanged. Thus the plasma resonances depend on the shape of the distribution, not at all on its scale (provided its extent is much less than a wavelength).

The question now is how many eigenvalues there are and what is their distribution for forms of $n_0(r)$ of interest. We approach an answer to these questions by considering stepped distributions approximating to (i) that inside a long discharge tube (Killiam 1930; Howe 1953), (ii) that in a meteor trail.

It is necessary first to show that resonant frequencies obtained from a stepped distribution approximating to a given smooth distribution will not differ significantly from those appropriate to the latter if the number of steps is large enough.

If K were everywhere >1 , as in ordinary dielectric materials, the internal field would certainly not be grossly affected by a transition from finely-stepped to smooth distribution. Thus one would only expect such effects (if any) in the present case where the mathematical features differ, namely, for eigenvalues crowded near $\lambda=1$ (see Section IV and Fig. 2) or arising from the behaviour of the solution near $K_1=0$, which latter we now consider.

If R in (5) is supposed integrated out from the origin from a zero value and arbitrary real slope at the origin, at $r=a$ where $K_1=0$ an imaginary part R^I is introduced into R proportional to $1/K'(a)$ (Kaiser and Closs 1952, equation (46)) but the real part R^R is the same at points A, B on either side of a such that $K_1(A)=-K_1(B)$ and both are near zero.

Thus the only effect of a finite $K'(a)$, as compared with the infinite $K'(a)$ with a stepped distribution, is on R^I . However, the condition of resonance is that $R^R \sim \text{const}/r$ and R^R in $r>a$ is quite independent of R^I if K_2 is negligible, since (5) is linear.

We may therefore calculate values of λ corresponding to resonance without considering R^I at all and for sufficiently many steps the stepped distribution should give the same answers as the smooth distribution to which it approximates.

IV. STEPPED DISTRIBUTIONS

In the regions $0, 1, 2, \dots, n$ (Fig. 1), $V_p = (B_p r + A_p r^{-1}) \cos \theta$, say, where $B_0=0, A_n=0$, and from the joining conditions at r_p ,

$$\begin{aligned} B_{p-1}r_p + A_{p-1}r_p^{-1} &= B_p r_p + A_p r_p^{-1}, \\ K_{p-1}(B_{p-1} - A_{p-1}r_p^{-2}) &= K_p(B_p - A_p r_p^{-2}). \end{aligned}$$

The condition for a non-zero solution is then

$$D_n \equiv \begin{vmatrix} r_1^{-1} & -r_1 & -r_1^{-1} & 0 & 0 & \dots & \dots \\ -r_1^{-2} & -K_1 & K_1 r_1^{-2} & 0 & 0 & \dots & \dots \\ 0 & r_2 & r_2^{-1} & -r_2 & -r_2^{-1} & 0 & \dots \\ 0 & K_1 & -K_1 r_2^{-2} & -K_2 & K_2 r_2^{-2} & 0 & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \dots & \dots & 0 & r_n & r_n^{-1} & -r_n \\ 0 & 0 & \dots & \dots & \dots & \dots & 0 & K_{n-1} & -K_{n-1} r_n^{-2} & -K_n \end{vmatrix} = 0. \quad (6)$$

Putting $K=1-\lambda f_r$, this is an equation of degree n to determine λ . This will have n roots, but only real roots and their distribution are here of interest, these giving the dipole mode frequencies. The case of two steps was considered by Kaiser and Closs (1952).

(a) *Discharge Tube Distribution*

The measured distribution of electron density across a discharge tube of cylindrical section given by Killian (1930) for a temperature of 18.6 °C can be fitted approximately by the parabolic curve

$$n_0=N_0\left\{1-0.6\left(\frac{r}{r_1}\right)^2\right\}, \dots\dots\dots (7)$$

where N_0 is the density at the centre and r_1 the radius of the tube.

We have taken stepped distributions approximating to this curve with successive values 2, 3, . . . 7 of n , the number of steps. The steps are at equal radial intervals and, except in the case $n=6$, the height of each is proportional to the value of (7) at its mid point, with $f=1$ at $r=0$. The values of λ for resonance and the corresponding values of K_p are shown in Table 1.*

It will be noticed that at least one of the K_p is negative (as required from Section III and as can be generally shown from the form of the determinant).

TABLE 1
VALUES OF λ FOR RESONANCE AND THE CORRESPONDING DIELECTRIC CONSTANTS K_p IN THE ANNULI BETWEEN STEPS, FOR VARIOUS NUMBERS n OF STEPS IN THE APPROXIMATING DISTRIBUTION AT RADII r_p

n	λ	Dielectric Constant, K_p						
		r_p	1	1/2				
2	1.226 2.684	f_p	0.6883	1.0				
			0.1560	-0.2262				
			-0.8477	-1.6844				
3	1.0888 1.4442 2.8846	r_p	1	2/3	1/3			
		f_p	0.5932	0.8644	1.0			
			0.3541	0.0588	-0.0888			
			0.1433	-0.2484	-0.4442			
			-0.7111	-1.4934	-1.8846			
4	1.048 1.2001 1.5973 2.9745	r_p	1	3/4	1/2	1/4		
		f_p	0.5457	0.7729	0.9243	1.0		
			0.4280	0.1900	0.0313	-0.0481		
			0.3451	0.0725	-0.1092	-0.2001		
			0.1283	-0.2345	-0.4764	-0.5973		
			-0.6233	-1.2989	-1.749	-1.974		

* General expressions for the coefficients of the powers of λ in (6) were found, but for six or more steps it was found necessary for accurate determination of the smaller eigenvalues to evaluate the determinant (using Crout's method) for neighbouring values of λ and interpolate to the eigenvalues.

TABLE 1 (*Continued*)

n	λ	Dielectric Constant, K_p							
		r_p	1	4/5	3/5	2/5	1/5		
5		f_p	0.5171	0.7103	0.8551	0.9517	1.0		
	1.0321		0.4663	0.2670	0.1174	0.01777	-0.03211		
	1.1202		0.3472	0.1034	-0.0795	-0.2014	-0.2624		
	1.3205		0.3172	0.06208	-0.1292	-0.2568	-0.3205		
	1.7102		0.1157	-0.2147	-0.4624	-0.6276	-0.7102		
	3.0217		-0.5625	-1.1462	-1.584	-1.876	-2.022		
6		r_p	1	5/6	2/3	1/2	1/3	1/6	
		f_p	0.5	0.66	0.8	0.9	0.96	1.0	
	1.215		0.3925	0.1981	0.0280	-0.0935	-0.1664	-0.2150	
	1.375		0.3125	0.0925	-0.100	-0.2375	-0.3200	-0.375	
	1.75		0.125	-0.155	-0.400	-0.575	-0.680	-0.75	
	3.061		-0.5305	-1.0203	-1.449	-1.755	-1.9386	-2.061	
7		r_p	1	6/7	5/7	4/7	3/7	2/7	1/7
		f_p	0.4841	0.6315	0.7543	0.8526	0.9263	0.9754	1.0
	1.027		0.5029	0.3515	0.2254	0.1245	0.0488	-0.0016	-0.0268
	1.146		0.4449	0.2760	0.1352	0.0225	-0.0620	-0.1183	-0.1465
	1.321		0.3605	0.1658	0.0035	-0.1262	-0.2236	-0.2885	-0.3209
	1.924		0.0685	-0.2151	-0.4514	-0.6405	-0.7823	-0.8768	-0.9241
	3.068		-0.4854	-0.9376	-1.3144	-1.6159	-1.842	-1.993	-2.068

The number of real eigenvalues is, for many steps, less than the number of steps and with increasing n the further modes introduced have values of λ crowding in near $\lambda=1$. For $n=7$ it was not feasible to determine accurately the eigenvalues closest to 1. Figure 2 shows the eigenvalues for increasing n ; while the dotted lines have no significance in detail (since the approximating stepped distributions are to a certain extent arbitrary) they do show a general trend.

In Figure 1 the lines of electric displacement \mathbf{D} corresponding to two values of λ are sketched and it is seen that in the mode with highest λ (i.e. highest frequency when the distribution is fixed, or highest density when the frequency is fixed and the density generally increased by raising the discharge current) the lines reach out furthest. This feature has been found generally in the modes studied and it shows that the highest λ mode will be that most strongly coupled to an external exciting field, hence that most strongly excited if losses are not greatly different for the neighbouring modes. However, the damping arising from energy loss near $K_1=0$ is least for the modes with largest K_1' there.

The experiments of Romell (1951) may be explained qualitatively by the foregoing. He found a number of resonant peaks in the reflection of transversely polarized waves from a cylindrical discharge tube as the discharge current was

varied and showed that their angular dependence was that of dipole, not multipole, modes. The highest peak was that with the greatest discharge current, corresponding to our greatest eigenvalue of λ . Quantitative comparison is not possible because the electron distribution across the tube was not measured by Romell.

Stepped electron distributions approximating to other shapes, e.g. spherical, may be similarly discussed with the same conclusion that there is in general a multiplicity of dipole modes, as there is also of multipole modes.

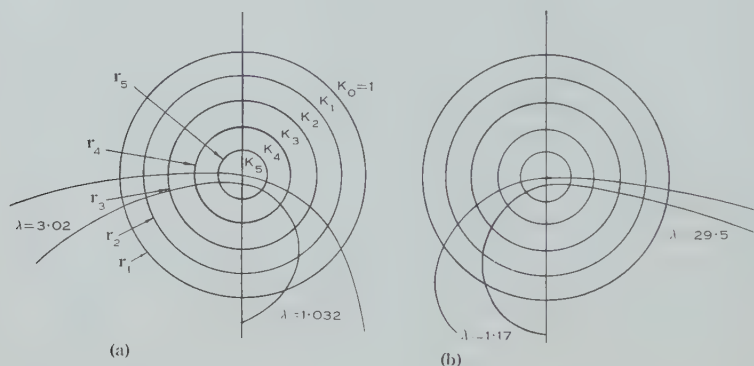


Fig. 1.—Lines of electric displacement in resonant modes for stepped distributions approximating to (a) that found in a discharge tube by Killian, (b) a Gaussian distribution. The left and right sides of each diagram correspond to the particular eigenvalues indicated.

(b) Gaussian Distribution

Taking $f(r) = \exp(-r^2)$ and approximating to this by five steps ($f_0, \dots, f_5 = 0, 0.0425, 0.1489, 0.4043, 0.7447, 1.0$; $r_1, \dots, r_5 = 1, 0.8, 0.6, 0.4, 0.2$) the equation for λ is

$$\lambda^5 - 44.619\lambda^4 + 515.084\lambda^3 - 2135.7\lambda^2 + 3475.9\lambda - 1888.7 = 0,$$

and the five real roots are: 1.17, 1.81, 3.44, 8.73, 29.5. The corresponding lines of induction are sketched in Figure 1 (b). Again the greater reach of the lines for the higher eigenvalues indicates stronger coupling to the exciting field. However, in considering an actual smooth Gaussian distribution, if the above values of λ are taken to give the resonant conditions, we note that at the radius where $K=0$ the slope of the curve for $K(r)$ is greatest for the eigenvalues 1.81, 3.44, consequently these modes will suffer least damping (see Section VI).

Kaiser and Closs discovered only one dipole mode by numerical integration with (in our notation) $\lambda = 2.4$. This corresponds presumably to our value 1.81. Their Figure 7 based on calculations with a smooth Gaussian distribution and numerical integration of (5) shows no evidence of more than one resonance, yet there is no reason to believe that the stepped distribution manufactures resonances with no counterparts for the smooth distribution. We can only

suggest that the intervals of λ (their f) chosen by those authors were not closely spaced enough to make the finer structure of the resonant behaviour apparent when the diffuseness damping (Section V) is present to the extent they assume.*

We would thus expect at least two maxima to be apparent in transversely polarized echoes from meteor trails (of diameter small enough for the quasi-static approximation) unless other conditions of the problem obscure their separateness.

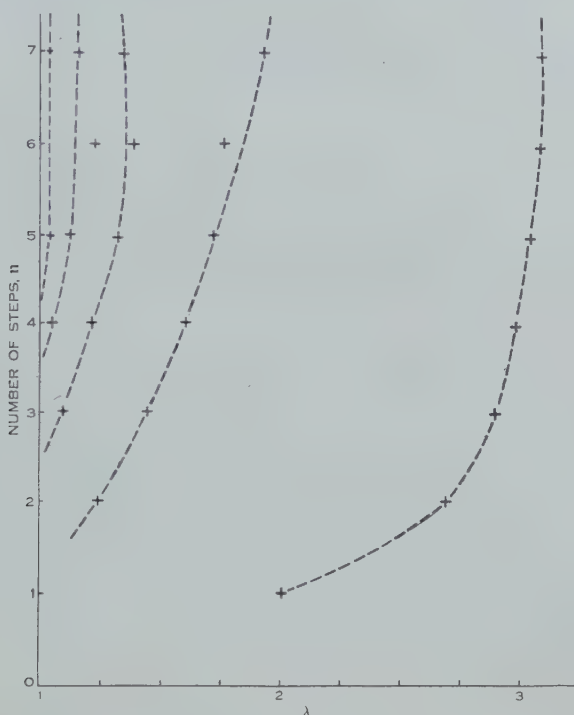


Fig. 2.—The values of λ for resonance as the number of steps is increased in a stepped distribution approximating to (7). A trend is suggested by the dotted lines.

(c) Nearly Uniform Cylinder

Some light on the distribution of the eigenvalues in general may be obtained as follows. We suppose the electron density distribution perturbed from a simple step, putting $f_r = 1 + \varepsilon g_r$, ε small, $r = 1, \dots, n$, so that in (6) $K_r = 1 - \lambda - \lambda \varepsilon g_r$. When $\varepsilon = 0$, $D_n(\lambda) = \frac{1}{2}(-2)^n(1 - \lambda)^{n-1}(2 - \lambda)(r_1 \dots r_n)^{-1}$, so that $D_n(\lambda)$ has an $(n - 1)$ -fold root at $\lambda = 1$, as well as the single root at $\lambda = 2$ corresponding to the

* *Note added in Proof.*—Dr. Kaiser has informed us that the intervals were quite close (namely, $\lambda = 0, 0.5, 1.0, 1.5, 2.0, 2.4, 2.5, 3.0, 4.0, 6.0, 8.0, 18.0, 20.0, 1000$). This suggests that the diffuseness damping was sufficient to prevent resolution of separate resonances. However, in the actual physical case the diffuseness damping may well be less than that calculated, see Section V.

eigenvalue $K = -1$ for a uniform cylinder. If we now consider terms in $\epsilon, \epsilon^2, \dots$ in succession it is seen by inspection of D_n that it is of the form

$$\begin{aligned} \frac{1}{2}(-2)^n(r_1 \dots r_n)^{-1}(1-\lambda)^{n-1}(2-\lambda) + b_1\epsilon\lambda(1-\lambda)^{n-2} + b_2\epsilon^2\lambda^2(1-\lambda)^{n-3} \\ + \dots + b_{n-1}\epsilon^{n-1}\lambda^{n-1} + b_n\epsilon^n\lambda^n, \end{aligned}$$

where the b_r do not contain the factor $(1-\lambda)$. Thus in the first order perturbation there is an $(n-2)$ -fold root at $\lambda=1$, one of the coincident roots having been split off and lying near 1 while the root near 2 is displaced slightly. Similarly each successive order of perturbation splits off a root from the remaining coincident roots at 1 and displaces those already split off. Not all these roots are necessarily real.

We thus expect, for a smooth distribution not too far removed from a simple step, a distribution of eigenvalues of λ in which there is one somewhere near 2, the remainder crowding up towards unity and perhaps becoming infinitely dense near unity (of course, in experimental observation, the latter would be obscured). It is apparent from Figure 2 that this general behaviour is there exemplified.

(d) Cavity Resonator Behaviour

If the distribution contains an annulus of finite width between r_p and r_{p+1} in which $K_p=0$, no lines of induction can penetrate this region and the oscillations (if any) inside the annulus are quite uncoupled with any taking place outside. The corresponding property of D_n is that it can then be expressed as the product of two determinants, the first containing only terms in $r_1, \dots, r_p, K_1, \dots, K_{p-1}$, the second only terms in $r_{p+1}, \dots, K_{p+1}, \dots$. For example, if there are two steps inside the annulus with $K_p=0$, the resonant frequency of oscillations inside the annulus is given by

$$K_p=0, \dots \dots \dots (8)$$

$$\frac{K_{p+1}}{K_{p+2}} = -\left(1 + \frac{r_{p+2}^2}{r_{p+1}^2}\right). \dots \dots \dots (9)$$

Regarded as equations to determine λ , (8) and (9) cannot in general both be satisfied, but given, say, f_p and f_{p+2} , there will be a certain value of f_{p+1} for which resonance is possible, namely,

$$f_p + (1 + r_{p+2}^2/r_{p+1}^2)(f_p - f_{p+2})/(1 - r_{p+2}^2/r_{p+1}^2).$$

Experimentally such a resonance might be sought with two concentric hollow cylinders of plasma, the central core having no plasma so that $K=1$ there. The average density of one plasma would need to be adjusted until resonance at some frequency near that expected was found. It is interesting to note that the dimensions of such a cavity resonator may be as small as we please in relation to a free-space wavelength.

V. ENERGY LOSS NEAR THE REGION WHERE $K=0$

It is well known that if one supposes energy losses, due, for example, to collisions, by the electrons to be adequately described by assigning a complex value $K_1 - jK_2$ to K there is a finite energy loss in a vanishingly thin region near where $K_1=0$ (and E therefore is very large) even though K_2 is vanishingly small (e.g. Makinson *et al.* 1951). If $K'_1(a)$ is the value of the gradient of K_1 where it vanishes and D is the amplitude of the (sinoidal) electric displacement, the rate of energy loss is readily seen to be proportional to

$$\frac{D^2}{K'_1(a)} \dots \dots \dots (10)$$

This loss is responsible for the damping of the resonances as discussed by Herlofson (1951) and Kaiser and Closs (1952).

Herlofson (1952) has pointed out that a more careful discussion of such results is needed and that a finite effect from vanishing collision frequency is not credible. We indicate below some limitations of the simple formal treatment. Any mechanism other than collisions which limits the amplitude of E at $K_1=0$ will remove or modify the energy loss derived as above.

Now firstly, one can describe the properties of a region containing free electrons by assigning to it a dielectric constant (real or complex) only as an approximation. As shown by Salpeter and Makinson (1949), if the velocities of the electrons (due to streaming or thermal motion) are large enough for their "transit time" (in travelling a distance over which E varies considerably) to be comparable with a period, such an assignment is not valid except as an approximation. In that approximation the values of K_1 and K_2 properly to be assigned to a point depend not only on the density and collision frequency at that point, but on the variation of E in the neighbourhood, which itself depends on the variation of K_1 and K_2 . Electrons may then acquire or lose energy where the field is strong and lose it elsewhere, (a) at the walls of a discharge tube, or (b) by collisions with gas molecules, or (c) by interaction with the electric field. Thus, if collision losses are very small, in the region near where K_1 vanishes E becomes very large and rapidly varying, and the proper value of K_2 is non-zero even if collision losses are supposed zero.

Mechanisms (a) and (b) lead to a net loss of energy from the field, while (c) does not, energy being merely transported from one part of the field to another by the electrons, giving a positive contribution to K_2 in some places, negative in others.

The energy loss near where $K_1=0$ will thus depend on the whole configuration in a complicated way but calculation might be attempted using the relations derived by Salpeter and Makinson (1949) in the special case where walls are remote, collisions are negligible, streaming is zero, and the electron temperature is supposed known and not too high.

One can however say that because at least of mechanism (c) the energy loss given by (10) is an overestimate and the "diffuseness" damping of resonances calculated on that basis is in some degree excessive.

A second reason for failure of (10) has been indicated by Herlofson (1952), namely, the neglect of second order effects arising from the very large values of E near $K_1=0$. For example, the amplitude of an electron motion will take it through regions of varying E . However, for sufficiently small applied field it would seem that the processes described in the preceding paragraphs must be more important.

VI. ACKNOWLEDGMENTS

One of us (R.E.B.M.) is indebted to Dr. T. R. Kaiser, Mr. R. L. Closs, Dr. N. Herlofson, and Professor H. S. W. Massey for discussions.

VII. REFERENCES

- HERLOFSON, N. (1951).—*Ark. Fys.* **3**: 247.
HERLOFSON, N. (1952).—U.R.S.I. Conference, Sydney.
HOWE, R. M. (1953).—*J. Appl. Phys.* **24**: 881.
KAISER, T. R., and CLOSS, R. L. (1952).—*Phil. Mag.* **43**: 1.
KILLIAN, T. J. (1930).—*Phys. Rev.* **35**: 1238.
MAKINSON, R. E. B., THONEMANN, P. C., KING, R. B., and RAMSAY J. V. (1951).—*Proc. Phys. Soc. Lond. B* **64**: 665.
ROMELL, D. (1951).—*Nature* **167**: 243.
SALPETER, E. E., and MAKINSON, R. E. B. (1949).—*Proc. Phys. Soc. Lond. B* **62**: 180.
TONKS, L. (1931).—*Phys. Rev.* **38**: 1219.

THE DECAY OF THE 7.68 MeV STATE IN ^{12}C

By R. G. UEBERGANG*

[Manuscript received February 9, 1954]

Summary

A polonium-beryllium source has been used to investigate the radiations in coincidence with the 4.43 MeV γ -ray leading to the ground state of ^{12}C . The coincident radiation consisted of a continuum due to neutrons and a γ -ray of energy 3.05 MeV and intensity 0.12–0.08 that of the 4.43 MeV γ -ray.

This indicates that the 7.68 MeV state of ^{12}C decays principally by cascade γ -emission.

I. INTRODUCTION

The second excited state of ^{12}C at 7.68 MeV has recently become the subject of considerable interest owing to the importance of the $^8\text{Be}(\alpha, \gamma)^{12}\text{C}$ reaction in those hot stars which have largely exhausted their central hydrogen. Under this condition the core is expected to contract and the central temperature to rise until thermonuclear reactions with the helium occur. Hoyle explains the original formation of elements heavier than helium by this process and concludes from the observed cosmic abundance ratios of $^{16}\text{O} : ^{12}\text{C} : ^4\text{He}$ that there should be a level in ^{12}C in the region of 7.6 MeV, that is, the $^8\text{Be} + \alpha$ reaction should have a resonance at about 0.25 MeV. Several reactions indicate the existence of a ^{12}C level in this region, the most recent and accurate measurement being from the $^{14}\text{N}(d, \alpha)^{12}\text{C}^*$ reaction which places the level at 7.68 ± 0.03 MeV above ^{12}C ground state (Dunbar *et al.* 1953). With the existence of the level confirmed it is important to know the relative probabilities of the different modes of decay which are energetically possible, namely,

- (i) decay to $^8\text{Be} + \alpha$,
- (ii) decay direct to ^{12}C ground state by γ -radiation or pair emission,
- (iii) cascade γ -decay to the ground state of ^{12}C .

Although the decay of the 4.43 MeV state of ^{12}C by γ -radiation is well known, no γ -radiation from the 7.68 MeV state to ground has been observed, and a lower intensity limit to the ratio of these two γ -rays has been given as 2500 : 1 (Beghian *et al.* 1953). The apparently forbidden nature of such a transition can be explained by both the 7.68 MeV state and ground state of ^{12}C having spin zero. Decay of this excited state by nuclear pair emission, however, has been reported by Harries and Davies (1952) who observed seven such pairs in cloud chamber measurements.

* Physics Department, University of Melbourne. (The sudden death of the author occurred in April.)

Beghian *et al.* (1953) have reported the existence of a 3.16 MeV γ -ray in the γ -ray spectrum from a polonium-beryllium source. Using a thick beryllium target, the intensity of the 3.16 MeV γ -ray was only 1 : 30 that of the 4.43 MeV γ -ray.

In the present paper the γ -ray cascade has been observed in the reaction ${}^9\text{Be}(\alpha, n){}^{12}\text{C}^*\gamma, \gamma{}^{12}\text{C}$. From the results it is possible to make an estimate of the relative probability of the decay of ${}^{12}\text{C}^*$ by this γ -ray cascade.

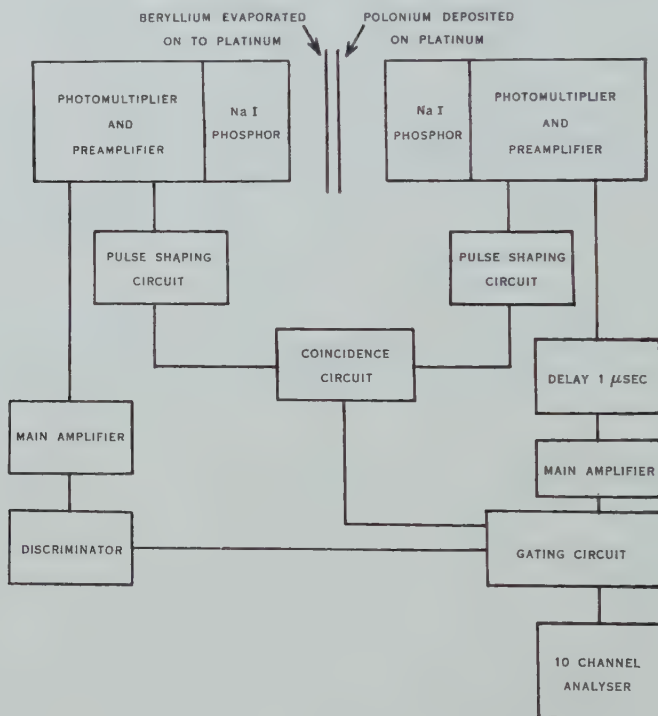


Fig. 1.—Block diagram of apparatus.

II. EXPERIMENTAL

The detection of a 4.43 MeV γ -quantum in one scintillation counter has been used to gate the output of a second scintillation counter so that the gated output consists essentially of pulses due to radiations in coincidence with the 4.43 MeV γ -ray. These radiations are :

- (a) neutrons (approx. 3 MeV) from ${}^{13}\text{C}^*$ to 7.68 MeV state in ${}^{12}\text{C}$,
- (b) γ -radiation (approx. 3 MeV) from 7.68 MeV state in ${}^{12}\text{C}$ to 4.43 MeV state,
- (c) neutrons (approx. 6 MeV) from ${}^{13}\text{C}^*$ to 4.43 MeV state in ${}^{12}\text{C}$.

The experiment therefore reduces to the problem of detecting radiation (b) in the presence of (a) and (c), where the relative intensities $a : b : c$ may be expected to be 1 : \leq 1 : 8 (Guier, Bertini, and Roberts 1952). This is achieved through the differential amplitude distribution of pulses from the detector

because the inelastic scattering of neutrons in the above energy region is expected to give rise to a continuum which is approximately logarithmic, decreasing to zero at an energy rather below the neutron energy. On the other hand γ -radiation at the above energy may be expected to give rise to peaks at energies corresponding to pair energy ($E_\gamma - 2m_0c^2$) and pair energy plus the capture of one annihilation quantum ($E_\gamma - m_0c^2$).

A block diagram of the detecting and analysing apparatus is shown in Figure 1. A 40 mc polonium-beryllium source was used, the polonium being deposited on a 1 cm diameter area on a platinum disk. Measurements were made with both thick and thin beryllium targets opposite the polonium and distant 0.020 in. from it. The thick target was a 0.010 in. thick beryllium disk while the thin target was made by evaporating beryllium on to a platinum disk to give a target approximately 200 keV thick.

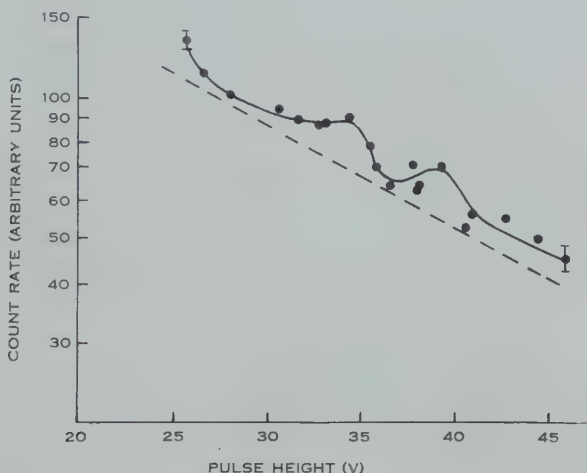


Fig. 2.—Typical pulse height distribution due to radiation in coincidence with 4.43 MeV γ -rays. Probable error has been indicated on two typical points.

Two scintillation counters (with sodium iodide crystal phosphors $1\frac{1}{2}$ in. diameter by 1 in. long) were used to detect the emissions. These counters were both placed as close as possible to the source. The output of a 0.2 μsec coincidence circuit between the two counters triggered a gate which passed on any pulses in counter *A* which were in coincidence with a 4.4 MeV γ -ray in counter *B*. These gated pulses were analysed on a 10 channel pulse amplitude analyser. The observed true to random ratio was 100 : 1 when using the thin target.

III. RESULTS AND DISCUSSION

Several measurements of the amplitude distribution of the gated pulses were made for each beryllium target. With the thick target the number of 6 MeV neutrons compared with the number of 3 MeV neutrons is estimated to be 30 : 1 while with the thin target the ratio is approximately 8 : 1. Therefore

any features of the coincidence pulse distribution which are emphasized in the thick target runs are due to neutrons, while any showing more distinctly in the thin target runs are due to γ -radiation.

The thick target runs give a smooth distribution with energy which is nearly exponential.

A typical pulse distribution due to radiation in coincidence with the 4.4 MeV γ -ray from a thin target is shown in Figure 2. The bumps in the curve which are evident at 34 and 39 V occurred in all thin target runs.

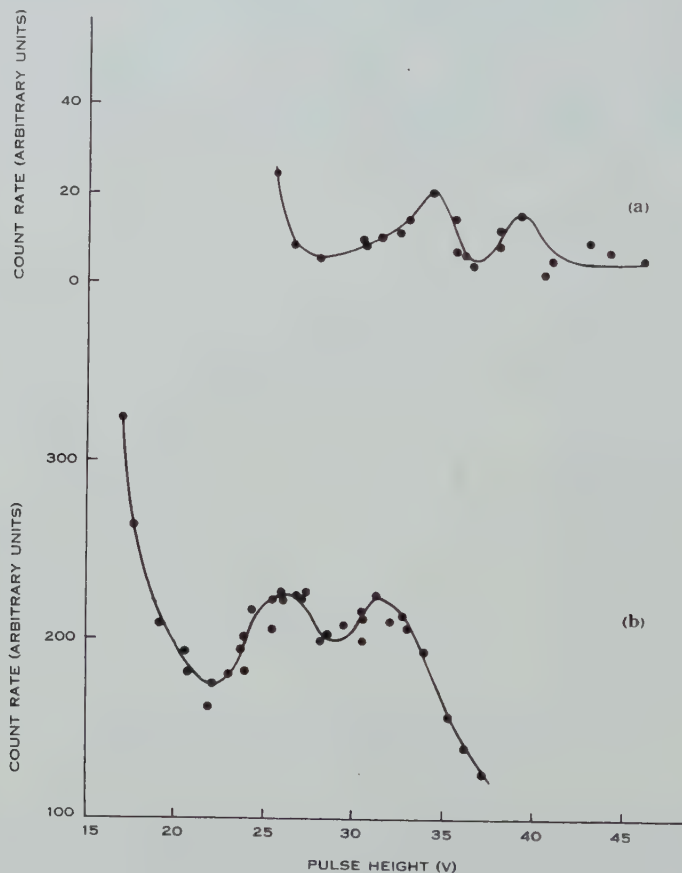


Fig. 3 (a).—Data shown in Figure 2 after background subtracted.

Fig. 3 (b).—Distribution of pulses due to 2.62 MeV γ -radiation from ThC'' .

When an exponential function similar to the thick target pulse distribution (represented by the straight line below the curve in Figure 2) is subtracted from the thin target measurement we obtain the curve shown in Figure 3 (a).

By comparison with the ThC'' pulse distribution (Fig. 3 (b)) the difference curve is seen to represent a single γ -ray line of energy approximately 3.0 MeV. Only the pair and pair-plus-one peaks appear because of the relatively poor resolution of the crystal used.

Using the ThC'' γ -radiation and the polonium-beryllium 4.43 MeV γ -ray pair peak for calibration the energy of the coincident γ -ray is estimated at 3.05 ± 0.1 MeV. This value of 0.1 MeV for the assigned error has been assessed from the repeatability of the pair peak in the several thin target runs and takes no account of errors in calibration. The energy, 3.05 MeV, found from the peak is lower than the measurement by scintillation counter pair spectrometer (Beghian *et al.* 1953) mentioned earlier and is also lower than the value required by the difference in energy of the two levels concerned. Because of uncertainties in the calibration, however, no significance is attached to this discrepancy.

From the peak to trough ratios in Figure 3 the fraction of counts due to 3.1 MeV γ -rays under the curve of Figure 2 may be estimated and from this figure and the geometry of the source and counters the ratio of 3.1 MeV γ -rays to 4.4 MeV γ -rays from the polonium-beryllium source has been calculated. This ratio lies between 0.08 and 0.12. Comparing these values with the photographic plate determination of the ratio of the number of neutrons leading to the excited states in ^{12}C as 1 : 8 (Guier, Bertini, and Roberts 1952), it appears that the decay of the state at 7.68 MeV is principally by cascade γ -emission.

IV. ACKNOWLEDGMENTS

The author wishes to thank Professor L. H. Martin for his continued interest and encouragement throughout the course of this work ; and Dr. D. N. F. Dunbar for advice and helpful discussions. Thanks are also due to Mr. N. W. Tanner who assisted in planning the experiment and with early measurements.

V. REFERENCES

- BEGHIAN, L. E., HALBAN, H. H., HUSAIN, T., and SANDERS, L. G. (1953).—*Phys. Rev.* **90** : 1129.
DUNBAR, D. N. F., PIXLEY, R. E., WENZEL, W. A., and WHALING, W. (1953).—*Phys. Rev.* **92** : 649.
GUIER, W. H., BERTINI, H. W., and ROBERTS, J. H. (1952).—*Phys. Rev.* **85** : 426.
HARRIES, G., and DAVIES, W. T. (1952).—*Proc. Phys. Soc. Lond. A* **65** : 564.

THE RESPONSE OF A SODIUM IODIDE SCINTILLATION COUNTER TO 18 MeV γ -RADIATION

By J. G. CAMPBELL* and A. J. F. BOYLE†

[Manuscript received February 18, 1954]

Summary

The probability distribution of ionization energy in a sodium iodide crystal 2.5 cm long by 2.5 cm diameter due to the absorption of 18 MeV quanta has been calculated taking account of side escape of electrons from the crystal due to multiple scattering. The results confirm the conclusions of the authors' earlier work.

I. INTRODUCTION

In a previous communication (Campbell and Boyle 1953, hereinafter referred to as C.B.) the authors discussed the response of sodium iodide scintillation counters to γ -radiation of energy up to 18 MeV. In view of the approximations employed in the theoretical section of that paper, one particular instance, the case of 18 MeV radiation in a crystal 2.5 cm long by 2.5 cm diameter, has been re-examined in greater detail.

II. METHOD

The Monte Carlo method described in C.B. was used, with the following modifications :

(i) Previously the secondary electrons were considered to travel in straight lines along the direction of the incident quanta, multiple scattering affecting only the distance travelled. Sideways deflexion is here taken into account, allowing escape through the side of the crystal.

(ii) The value used in C.B. for the ionization loss of fast electrons in sodium iodide is too low.‡ The value now used is 5 MeV/cm, which, from the data for sodium iodide given recently by Sternheimer (1952), holds within 10 per cent. for electrons in the energy range 4–20 MeV.

(iii) In C.B., the probability of reabsorption of isotropically emitted bremsstrahlung was calculated on the approximating assumption that it all originated at the central point of the crystal. For the present work, the published data on self-absorption of large sources (Dixon 1951) were used to derive this probability for a uniform distribution of origin throughout the volume.

* Physics Department, University of Melbourne.

† Research School of Physical Sciences, Australian National University, Canberra.

‡ The authors are indebted to Dr. G. W. Hutchinson, of the University of Glasgow, for drawing their attention to this error.

(iv) The method of calculation does not allow an electron to emit more than one bremsstrahlung photon in any one section of length. In C.B., these sections were chosen to be 0.25 cm, whereas they are here 0.10 cm. The double emission thus neglected is reduced in proportion to the square of the ratio of these lengths.

(v) The lowest energy bremsstrahlung photons considered are 0.25 MeV, instead of 0.5 MeV as in C.B. Although the energy content of each such photon is small, they occur frequently.

A set of life histories of individual secondary electrons, with each of the initial energies 3, 6, 9, 12, 15, and 18 MeV, was compiled, without regard to deviations of path. Then, to introduce sideways deflexion, a formula given by Janossy (1948, p. 319) was used. This states that the mean square displacement of an electron passing through a homogeneous absorber of thickness z is

$$x^2 = E_s^2 \int_0^z \frac{z'^2 dz'}{E(z')^2},$$

where $E(z')$ is its energy in MeV at depth $(z - z')$ and E_s is 21 MeV, distance being measured in radiation lengths. The cross section of the crystal was divided into five concentric annuli of equal area, and the electrons were assumed to start in each of these in turn. By a graphical method similar to one since published by Dickinson and Dodder (1953), the fractions of electrons escaping through the side of the crystal were determined for each section of length, for each initial energy and each annulus of cross section. These results were applied to the life histories. The length of the crystal was resubdivided into sections of 0.25 cm, and the ionization distribution was obtained for each case.

From these 300 distributions for the six initial energies, it was necessary to find those for initial energies from 0 to 18 MeV in steps of 0.5 MeV. To carry out this interpolation it was necessary to convert the distributions, which were in the form of histograms subject to statistical fluctuation, into an analytical form. The form which was arrived at for the distribution in energy E of the ionization in the crystal by an electron of initial energy E_0 from a given annulus having travelled a given distance through the crystal is

$$A\delta(E - E_1) + (1 - A)E^{p-1}(E_0 - E)^{q-1}/E_0^{p+q-1}B(p, q),$$

where $\delta(E - E_1)$ is the delta function and $B(p, q)$ the beta function. The first term is the contribution of those electrons whose passage through the crystal has been so far uneventful, so that they have all lost the same amount of energy E_1 by ionization. The value of A was determined directly from the peak in each histogram. The second term is due to those electrons which in addition have emitted bremsstrahlung or have escaped through the side of the crystal. The parameters p and q were found from the first and second moments of each histogram after the peak had been removed. The form of this term, which is known as the beta distribution, was found in each case to approximate well to the shape of the Monte Carlo histogram.

The ionization distributions for the six initial energies having thus been converted to analytical form, those for the intermediate energies were derived by linear interpolation of the three parameters A , p , and q . The 1800 distributions so obtained were tabulated numerically.

From these single-electron distributions, the distributions for pairs of electrons with initial kinetic energies totalling 17 MeV were found. This integration of tabulated data, which involved hundreds of thousands of multiplications, was carried out on "Hollerith" punched-card equipment, through the cooperation of the Commonwealth Bureau of Census and Statistics and of the Australian Wool Realization Commission. The pair distributions were then weighted with their respective probabilities, treated for capture of annihilation radiation using Figure 4 of C.B., and combined, to give the distribution of energy lost by ionization in the crystal by quanta initially absorbed by the pair production process.

The distribution due to quanta absorbed by the Compton process (only 20 per cent. of the total at this energy) was calculated using representative electron distributions. When combined with the pair-production distribution, the final calculated ionization distribution was obtained for 18 MeV quanta.

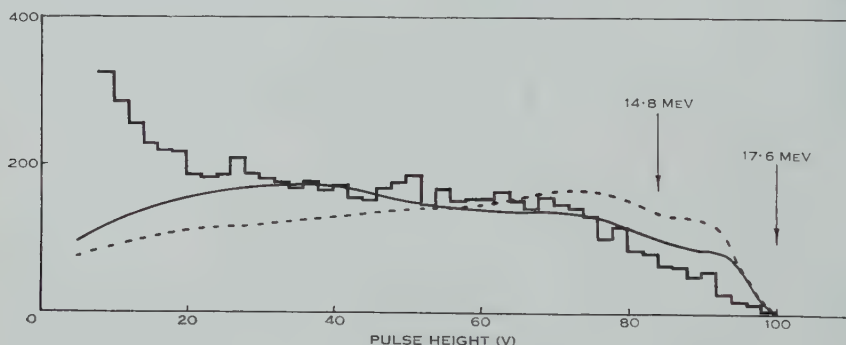


Fig. 1.—Results for ${}^7\text{Li}(p,\gamma)$ radiation.

Histogram: experimental, uncollimated.

Continuous curve: theoretical, uncollimated.

Dotted curve: theoretical, collimated.

III. RESULTS AND DISCUSSION

For a comparison with experiment, the authors' previously published results for the ${}^7\text{Li}(p,\gamma)$ radiation were used. The experimental pulse-height distribution, taken from Figure 10 (a) of C.B., is reproduced as the histogram in Figure 1. The arrows indicate the energy calibration arrived at using the ${}^{19}\text{F}(p,\alpha\gamma)$ radiation.

In determining the theoretical distribution for this radiation, the component at 14.8 MeV was assumed to give a distribution of the same shape as the more intense one at 17.6 MeV, except that its high energy edge was spread by the natural line breadth of 2 MeV. The further spreading due to statistical effects in the photomultiplier was ignored, being less than 2 per cent. at the high energy

end. The resulting theoretical distribution, scaled horizontally to fit the energy calibration and vertically to fit the histogram, is shown by the smooth continuous curve in the figure. Apart from the usual low energy divergence, the agreement is satisfactory.

The effect of sideways loss of electrons by multiple scattering could be minimized by collimating the beam of γ -radiation to a narrow pencil centred on the crystal axis, which, however, would have little effect on bremsstrahlung loss. The distribution to be expected in the case of a pencil 1.12 cm in diameter (corresponding to the centremost of the five annuli considered) is shown by the dotted line in the figure. This was not tested experimentally, since the resolution is not significantly improved.

The results of this more thorough calculation confirm the conclusions of C.B., that crystals of dimensions of an inch or two are too small for spectrometry of γ -radiation of this energy, and that the most important limiting factor is the escape of bremsstrahlung photons. The improvement possible with larger crystals has since been shown by Foote and Koch (1954), who have clearly resolved the two components of the ${}^7\text{Li}(p, \gamma)$ radiation using a sodium iodide crystal 8 in long by 5 in diameter.

IV. ACKNOWLEDGMENTS

Valuable assistance was received from Mr. J. Anderson, of the Commonwealth Bureau of Census and Statistics, who planned the punched-card programme, and from Mr. A. Wilson, of the Australian Wool Realization Commission, who carried it out.

The authors have appreciated the continued interest of Professor L. H. Martin.

V. REFERENCES

- CAMPBELL, J. G., and BOYLE, A. J. F. (1953).—*Aust. J. Phys.* **6**: 171.
DICKINSON, W. C., and DODDER, D. C. (1953).—*Rev. Sci. Instrum.* **24**: 428.
DIXON, W. R. (1951).—*Nucleonics* **8** (4): 68.
FOOTE, R. S., and KOCH, H. W. (1954).—*Rev. Sci. Instrum.* **25** (in press).
JANOSSY, L. (1948).—"Cosmic Rays." (Oxford Univ. Press.)
STERNHEIMER, R. M. (1952).—*Phys. Rev.* **88**: 851.

THE MEASUREMENT AND REDUCTION OF DISTORTION IN THICK EMULSIONS

By V. D. HOPPER,* Y. K. LIM,* and MADELINE C. WALTERS*

[*Manuscript received March 1, 1954*]

Summary

A study has been made of some of the factors influencing the distortion of developed images of tracks in Ilford G5 400 μ emulsions. It has been found that the effect of distortion on scattering measurements can be reduced to a negligible amount by developing the emulsions at 15 °C and drying with an alcohol-water solution of increasing concentration. A simple quantitative description of the distortion can be obtained by fitting a sinusoidal series to the curve of a high energy track traversing the emulsion.

I. INTRODUCTION

The use of thick emulsions is of considerable advantage in cosmic radiation studies, since the possibility that an event will be completely contained within a single layer is thereby greatly increased. Emulsions having a thickness of several millimetres have been employed but these are seriously distorted during processing. Also, in order to study an event in detail, a high power objective of $95\times$ is used and, as this has an effective working distance of about 250 μ , emulsions which in their final developed state have thicknesses greater than this cannot be conveniently studied. When an Ilford G5 nuclear emulsion plate of thickness 400 μ is developed and dried, the final thickness of emulsion is about 150 μ , and a 600 μ plate is reduced to 225 μ . Thus emulsions of thickness 600 or 400 μ are generally employed in cosmic ray studies and layers of such emulsions are used if greater effective thickness of emulsion is required.

When a nuclear emulsion plate of, say, 400 μ thickness on glass backing is developed, it also suffers a small general contraction in the planes parallel to the plate. The contractions are not perfectly uniform with the result that a straight track in the undeveloped emulsion shows curvature in the developed emulsion, the curvature becoming more pronounced as the inclination of the track to the plane of the emulsion-glass surface increases. With the nuclear emulsion technique the energy of a charged particle can be determined from a measurement of scattering, but, in order to accurately determine scattering effects, the distortion must be reduced to a small amount. In the measurement of scattering, additional sources of error enter due to the mechanical irregularities of the microscope stage, and observational errors. The object of this work has been to investigate factors which influence distortion and to find a method of development which produces clearly visible minimum ionizing tracks and also reduces the error due to distortion below other errors.

* Physics Department, University of Melbourne.

II. MEASUREMENT OF EMULSION DISTORTION

It may be assumed that over a few millimetres of a plate needed to measure the scattering of a track, the curvature distortion is due to lateral displacement of parallel layers at different depths relative to the glass backing. The lateral displacement Δ of a point at a height Z in a distorted emulsion of thickness d , may be expressed as a function of Z/d by the power series

$$\Delta = K_1(Z/d) + K_2(Z/d)^2 + \dots \quad (1)$$

proposed by Cosyns and Vanderhaeghe (1950). The first term in the series represents a change in the orientation of the tracks in the emulsion and this does not affect scattering measurements while second and higher order terms give rise to a curvature. A rapid method of measuring the second coefficient K_2 in the series has been proposed by Major (1952) who showed that this coefficient is four times the distance δ between the mid point of the chord joining the ends of the track and the point of intersection of the track with the plane $Z=d/2$. In practice the centre of the eyepiece scale is placed at the centre of the chord and the scale is then rotated until it is directed at the point of the track in focus at half-depth, when δ may be read off. The assumption is made that third and higher order terms in Z/d are zero, or may be neglected. This assumption is not always justified, S-shaped distortion being frequently observed and in these cases K_3 is greater than $0.5 K_2$. Even for tracks which appear C-shaped the value of K_3 may be appreciable.

The lateral displacement Δ , resulting from distortion, of a point in an originally straight track is more conveniently expressed by the series of angular functions.

$$\Delta = K_1(Z/d) + a \sin \pi(Z/d) + b \sin 2\pi(Z/d) + \dots \quad (2)$$

The number of terms involved in a particular distortion is then apparent from the number of points of inflection along the track. The advantage in this method of expression is that the coefficients a and b are readily determined from two simple measurements on a single track, while higher order distortions are seldom observed. The parameter δ , whose measurement is described above, is equal to the coefficient a . If S-distortion is appreciable compared with C-distortion (i.e. $b > a/2$), b may be determined by an observation of the height Z in the emulsion at which the track intersects the chord joining its ends, since at this point

$$a \sin \pi(Z_1/d) = -b \sin 2\pi(Z_1/d). \dots \quad (3)$$

Neither of these measurements is affected by the inclination of the plane containing the track with the direction of δ . Any high energy track in the region investigated may therefore be selected for measurement.

In order to find out how closely the form of the C-shape distorted tracks in an emulsion fitted a first order distortion of the type $\Delta = a \sin \pi(Z/d)$, the projection curves on a horizontal plane, of 10 steeply inclined tracks were measured to an accuracy of $\pm 1 \mu$. The tracks chosen for measurement lay at angles between 18 and 90° with the direction of distortion. From plots of these

projections the horizontal distance between chord and track, in the direction of the distortion, δ , could be determined at various points along the track. Values of Δ/δ plotted as a function of depth may be seen from Figure 1 to approximate to a sine curve within limits of ± 5 per cent.

III. THE APPARENT MEAN SCATTERING ANGLE OF A DISTORTED TRACK

The apparent mean scattering angle which will be obtained from direct measurements on a track in a distorted emulsion will depend upon the orientation of the track, and will be a maximum for tracks lying in a plane perpendicular to the direction of distortion. The value of the mean scattering angle resulting from distortion may be estimated from measured values of the coefficients a and b of equation (2).

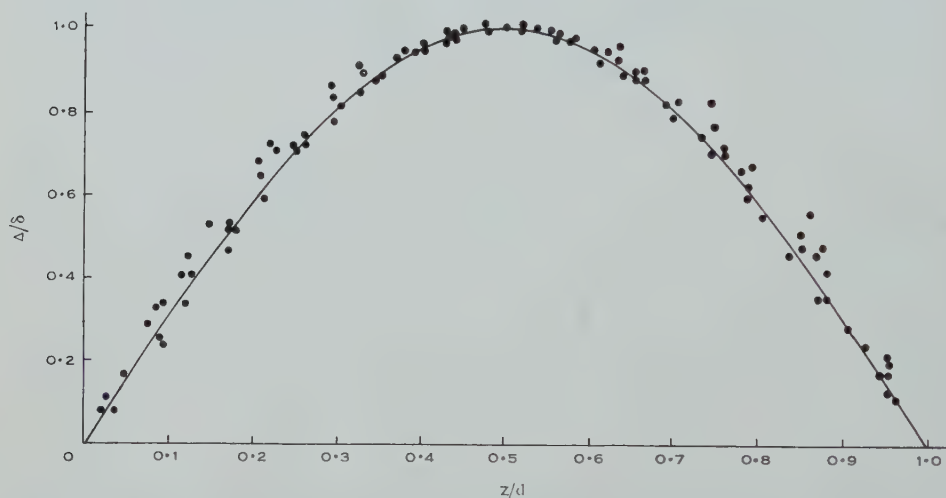


Fig. 1.—Lateral displacement in the direction of distortion as a function of depth for plate showing C-distortion.

For the case of a first order distortion, when $b=0$, the x and y coordinates of the projection on a horizontal plane of a straight track, lying at an angle θ to the direction of distortion and extending through the whole depth of the emulsion, may be approximately represented by the equation

$$y = a \sin \theta \sin \pi x/l_0, \quad \dots \dots \dots (4)$$

where l_0 is the projected length of the chord joining the ends of the track. The mean scattering angle $\bar{\alpha}_d$ observed for such a track is then given by

$$\bar{\alpha}_d = \frac{a\pi c}{l^2} \left(\frac{Z_2 - Z_1}{d} \right) \left\{ \cos \left(\frac{\pi Z_1}{d} \right) - \cos \left(\frac{\pi Z_2}{d} \right) \right\} \sin \theta, \quad \dots \dots \dots (5)$$

where l is the projected length of the track extending from a depth Z_1 to Z_2 in the emulsion, and c is the cell length measured in the direction of l .

Numerical values of $\bar{\alpha}_d$ have been estimated for tracks lying in a plane perpendicular to the direction of distortion, and are given in Figure 2. These values represent the maximum error which could result from distortion for any particular dip angle.

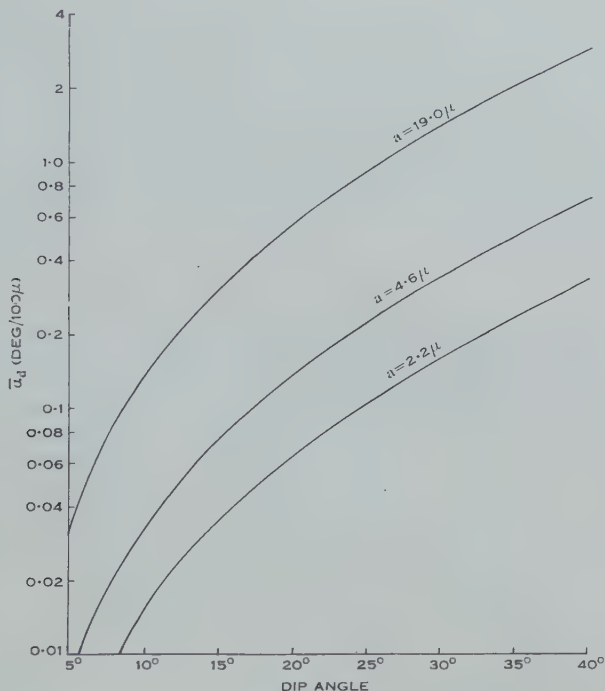


Fig. 2.—Maximum mean scattering angle due to distortion for the case of 100 μ cell length plotted against dip angle in unprocessed plate for three values of a .

IV. AN INVESTIGATION OF THE CURVES OF EMULSION DISTORTION

In order to obtain uniform development throughout a thick emulsion layer the "temperature development method" of Dilworth, Occhialini, and Payne (1948) may be employed. During the warm stage of the process the emulsion tends to lose its elasticity and may flow slightly, and when the emulsion is later cooled such distortions become frozen in. These distortions can be reduced by lowering the temperature of the warm stage, though this tends to alter the characteristics of the development and results in an increased fog level which renders track recognition difficult (Herz 1953). In order to find the optimum warm stage temperature a series of developments has been carried out at different temperatures and the degree of distortion produced measured in each case.

The other principal cause of distortion is uneven drying of the processed emulsion. Since nuclear emulsions contain an abnormally high concentration of silver bromide, the emulsion thickness is reduced to less than half its original value during processing. The contraction occurs mainly in the drying stage,

and during washing the emulsion absorbs water and becomes swollen to a greater thickness than at any other stage in the processing. If drying should be uneven, as is the case when the plate is dried rapidly in air, large distortions occur at the edge as has been illustrated by Major (1952). Several different methods of drying have therefore been tried out and were found to have a considerable influence on the amount of distortion.

Other possible sources of distortion which have been suggested are rapid changes of temperature before and after the warm stage, and sudden changes in pH of the processing solutions. These have not been studied in the present investigation. It has also been suggested by Bonetti, Dilworth, and Occhialini (1951) that distortion produced in the manufacture of the plates may be released during development. If this is in fact the case the developed emulsion will contain some distortions whatever precautions are taken during processing.

V. EFFECT OF WARM STAGE TEMPERATURE ON DISTORTION

Ilford G5 plates, 400 μ thick, were developed in an amidol-bisulphite developer of the type described by Dainton, Gattiker, and Lock (1951). The plates were presoaked in distilled water at 4 °C for 2 hr. After soaking in cold developer for 2 hr at 4 °C, the surface developer was removed and the plates then transferred to a hot plate at temperatures of 27, 20, and 15 °C for appropriate times. Fixation and washing were carried out at temperatures below 5 °C. The plates were immersed in a 1 per cent. solution of glycerine in water for 2 hr before being dried to prevent subsequent stripping of the emulsion from the glass backing. In order to dry the plates uniformly they were immersed for 2 hr at room temperature in each of a series of alcohol-water solutions, in which the alcohol concentration was increased in four stages up to 80 per cent. The plates were then allowed to dry in a horizontal position in a wooden box. Distortion measurements were made at 1 cm intervals over the entire area of the plates and the mean values of the distortion coefficients a and b obtained. The results are shown in Table 1.

TABLE 1
DISTORTION MEASUREMENTS RELATED TO TEMPERATURE OF DEVELOPMENT

Plate No.	Development at Warm Stage		Mean Distortion Coefficients		
	Temp. (°C)	Time (min)	a (μ)	b (μ)	$g_{\min.}^*$
1	27	23	19.0 ± 1.2	1.54 ± 0.15	14.8 ± 0.4
2	20	65	17.8 ± 0.5	0.0	12.7 ± 0.3
3	15	120	4.6 ± 0.3	1.90 ± 0.20	10.9 ± 0.3
4	26	30	12.5 ± 0.2	—	—
5	15	120	3.3 ± 0.6	0.31 ± 0.03	—
6	7	180	4.0 ± 0.6	0.31 ± 0.03	11.0 ± 0.3

* $g_{\min.}$ represents the number of grains for 50 μ for minimum ionizing particle.

Plates 4 and 5 were developed by the same method as plates 1-3 except that the drying process was carried out at a temperature below 5 °C. They show the same general reduction in distortion with decrease in the temperature of the hot stage.

An attempt was made to reduce the distortion still further by processing a plate at a constant temperature of 7 °C throughout, as discussed by Edgar and Herz (1953), their method being modified in this case by using the alcohol-water drying method. Temperature fluctuation of $\pm 1^\circ$ occurred about the mean temperature of 7 °C during the processing. Measurement of the distortion coefficients at about 40 evenly spaced points on one such plate yielded the comparatively low values quoted in Table 1. This method of development was not, however, considered as satisfactory as that of the plates developed at 15 °C, since the fog density in the processed emulsion particularly near the surface was abnormally high.

VI. DISTORTION PRODUCED DURING DRYING

In order to investigate the distortion which results from different methods of drying, four 400 μ thick G5 plates were developed as described in the previous section, at a warm stage temperature of 15°, and then dried by different methods. Plates 7 and 8 were dried in air immediately after washing, plate 8 being left in an enclosed space maintained at high relative humidity by the presence of water, while plate 7 was allowed to dry in air of low humidity. Plates 9 and 10 were soaked for a period of 2 hr in each of a series of alcohol-water solutions in which the alcohol concentration of successive baths was increased by steps of 15 per cent. After immersion in the 60 per cent. alcohol solution plate 9 was dried in air, while plate 10 was taken through the whole series and finally immersed in absolute alcohol for 2 hr before being allowed to dry. Measurements of the distortion coefficient α were made at regular intervals over each plate and values corresponding to points more than 1 cm from the edges are quoted in Table 2.

TABLE 2
DISTORTION MEASUREMENTS RELATED TO METHOD OF DRYING

Plate No.	Method of Drying	Distortion Coefficient α (μ)
7	Air (evaporation, low humidity)	10.96 ± 0.90
8	Air (evaporation, high humidity)	10.45 ± 0.57
9	Dehydration in alcohol solutions to 60%	6.37 ± 0.30
10	Dehydration in alcohol solutions to 100%	2.15 ± 0.17

The distribution of distortion on plates 10 and 7 is shown in Figures 3 and 4. These illustrate the difference between a plate dried in air, when drying proceeds from the edges, producing a high degree of distortion near the edges and less at the centre, and a plate from which the water has been removed uniformly

with alcohol, which shows a much more even distribution of distortion. It is interesting to note that the longer drying time necessary to remove the excess water in a humid atmosphere compared with the time required in a dry atmosphere has had little effect on the degree of distortion produced.

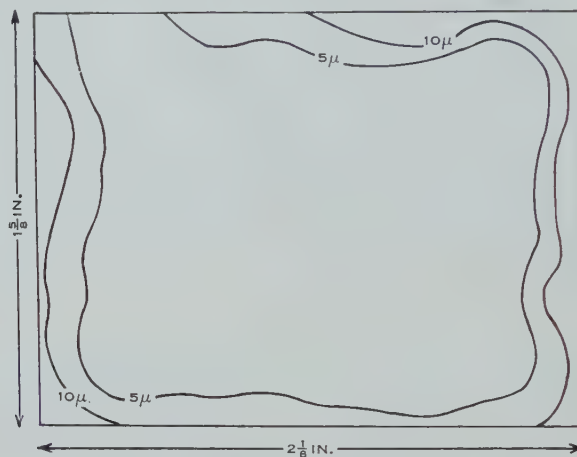


Fig. 3.—Contour map of distortion δ for plate 10.
Emulsion dehydrated with alcohol.

VII. COMPARISON OF DISTORTION AND NOISE LEVEL ERRORS

The minimum distortion value of a so far produced, 2.2μ in a 400μ thick emulsion, is satisfactory from the point of view of scattering measurements,

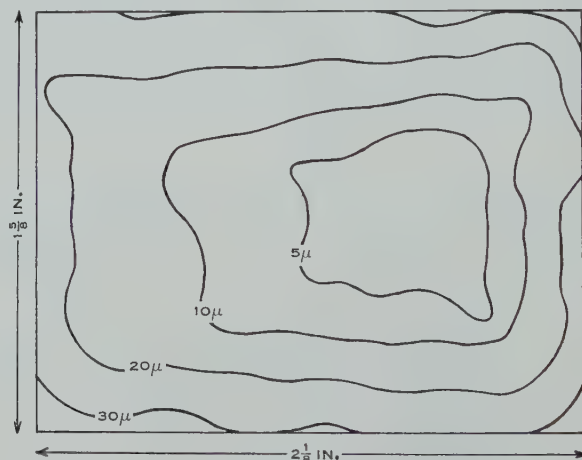


Fig. 4.—Contour map of distortion δ for plate 7.
Emulsion dried in air.

since the error arising from such distortion is small in comparison with the noise level due to mechanical irregularities of the microscope stage and observational errors. If tracks which traverse an emulsion free of distortion are measured, the

accuracy of the measurement of scattering depends on the length of track in the emulsion. For steep tracks few successive cell lengths can be employed and these must be comparatively short, resulting in low statistical accuracy. The noise levels in degrees per 100 μ for varying cell lengths have been measured for the microscope employed in scattering measurements. These are compared in Table 3 with the maximum distortion errors in degrees per 100 μ , $\alpha_{d(\max.)}$, calculated for the same cell lengths for tracks having 17 successive cell lengths in the distorted emulsion which has a value of a of 2.2 μ . It will be observed that, for the track of cell length 100 μ and 1700 μ length in the emulsion, the error in the scattering measurement is only one-fifth of the error due to the noise level and this ratio decreases with increasing cell length and corresponding length of track in the emulsion. The lower limit to the small-angle scattering which is measurable thus depends primarily on the noise level and it is not necessary to make any correction to the scattering measurement for distortion for such plates.

TABLE 3
NOISE LEVEL ERRORS COMPARED WITH DISTORTION ERRORS (PLATE 10)

Cell Length (μ)	Corresponding Track Length (μ)	Maximum Dip Angle (deg)	Maximum Distortion (deg/100 μ)	Noise Level (deg/100 μ)
100	1700	13.3	0.028	0.15
200	3400	6.7	0.0095	0.081
400	6800	3.4	0.0033	0.036

VIII. CONCLUSIONS

These investigations have shown that the two principal causes of emulsion distortion are (1) the increased plasticity and consequent flow of the emulsion during the warm stage of development, and (2) the uneven contractions produced by non-uniform drying. Distortion arising from the first of these causes may be minimized by reduction in the temperature of the warm stage, though the change in the development action which then occurs sets a lower limit to this temperature. A warm stage temperature of 15 °C appears to result in satisfactory development while reducing distortion to a low level. Distortion produced in the drying stage can be minimized by gradual and uniform dehydration of the emulsion in alcohol-water solutions of increasing concentration, and finally in absolute alcohol. For plates developed in this manner the error in scattering measurements due to distortion is very much smaller than that due to noise level.

IX. ACKNOWLEDGMENTS

The authors gratefully acknowledge the continual support and interest of Professor L. H. Martin. One of the authors (M.C.W.) is assisted by the British Memorial Fund.

X. REFERENCES

- BONETTI, A., DILWORTH, C. C., and OCCHIALINI, G. P. S. (1951).—Bull. Cent. Phys. Nucléaire Bruxelles No. 13b.
- COSYNS, M. G. E., and VANDERHAEGHE, G. (1950).—Bull. Cent. Phys. Nucléaire Bruxelles No. 15.
- DAINTON, A. D., GATTIKER, A. R., and LOCK, W. O. (1951).—*Phil. Mag.* **42** : 396.
- DILWORTH, C. C., OCCHIALINI, G. P. S., and PAYNE, R. M. (1948).—*Nature* **162** : 102.
- EDGAR, M., and HERZ, A. J. (1953).—*Proc. Phys. Soc. Lond. A* **66** : 115.
- HERZ, A. J. (1953).—*J. Sci. Instrum.* **29** : 60.
- MAJOR, J. V. (1952).—*Brit. J. Appl. Phys.* **3** : 309.

OBSERVATIONS OF THE 21 CM LINE FROM THE MAGELLANIC CLOUDS

By F. J. KERR,* J. V. HINDMAN,* and B. J. ROBINSON*

[*Manuscript received February 22, 1954*]

Summary

A survey has been made of the 21 cm line from neutral interstellar hydrogen in the Magellanic Clouds. These are the first observations of this radiation from an extragalactic source.

The observations show that neutral hydrogen extends well beyond the easily visible regions of each Cloud, the gas being in each case less concentrated towards the nucleus than are the bright stars. From the total radiation received, the masses of neutral hydrogen are calculated to be 6×10^8 and 4×10^8 solar masses for the Large and Small Cloud respectively. Since the Small Cloud is believed to contain much less dust, this means that the ratio of gas to dust is very different in the two systems.

The measurement of Doppler frequency shifts has yielded extensive new information about radial velocities within the Clouds. The velocity distribution indicates that both Clouds are rotating. Detailed discussion of the velocities has been deferred for subsequent publication.

I. INTRODUCTION

The prediction (van de Hulst 1945) and subsequent detection (Ewen and Purcell 1951) of a radio-frequency line in the spectrum of interstellar hydrogen have opened up new possibilities in astronomical exploration. Extensive observations of 21 cm line radiation from the Galaxy have been reported by Christiansen and Hindman (1952) and van de Hulst (1953). This paper describes the first observations of the line from an extragalactic source.

Besides its immediate value in revealing the existence and some of the physical properties of the interstellar gas, this line radiation shares with the continuous spectrum radio-frequency radiation the ability to penetrate the clouds of interstellar dust which severely limit visibility in optical astronomy. In addition, it has the important advantage that it is a discrete line, so that recognizable frequency displacements are produced by the motions of the emitting gas. It thus combines for the first time in astronomy the "seeing power" of the radio wave with the possibility of measuring the (radial) velocities of the source regions. Hence both the motions and the distribution of gas in a large volume of space can be studied.

The radio observations do not by themselves provide sufficient information to deduce three-dimensional structure, because distance cannot be measured directly. It can in some cases, however, be inferred from other evidence.

* Division of Radiophysics, C.S.I.R.O., University Grounds, Sydney.

For instance, the systematic nature of galactic rotation effects has been used for this purpose by the Leiden group (van de Hulst 1953) to locate portion of the spiral structure of the Galaxy. For an external galaxy, there is the important simplification that all parts of the system are at substantially the same distance. Thus a low-resolution two-dimensional picture can be derived directly, and to some extent the details of the line profile can provide information about the distribution of the gas in depth.

The Magellanic Clouds, the nearest extragalactic systems, are specially suitable for study. They are sufficiently far away to be viewed as isolated systems, and yet are sufficiently near for their internal structure to be discerned in some detail. Studies of the Clouds have been very important in the development of optical astronomy.

An exploratory survey has been made over the region of the sky containing the Clouds.* The 21 cm line has been detected from each Cloud over a large area extending well beyond its previously known limits. Values have been derived for the total mass of neutral hydrogen in each Cloud, and information obtained about the two-dimensional distributions of intensity and velocity. The preliminary results indicate that observations of this line will be able to provide a detailed picture of the structure and dynamics of the Clouds, leading to a number of important inferences concerning galaxies of this type.

The present results are preliminary in character. They were obtained with a new type of receiver at present under development, and it should be possible to carry out a more complete and precise survey before long. Consequently, the scope of this paper is limited to the presentation of broad overall results, together with some discussion of the type of information which could be derived from more extensive data.

II. RECEIVING TECHNIQUE

(a) *Equipment*

The energy received from any given direction is dispersed over a finite frequency band (typically $\frac{1}{4}$ – $\frac{1}{2}$ Mc/s), owing to random and systematic motions of the gas. Thus both the frequency distribution and intensity of line radiation vary across the sky.

In the usual method of observing the line, the receiver frequency is swept slowly over an appropriate band. At the same time, the receiver is switched rapidly between two adjacent frequencies. This produces a modulated signal which is independent of the background radiation, and minimizes changes of zero level arising from gain variations. Since the intensity is very low, the output must be integrated over a period of time to reduce the noise fluctuations. In consequence, the frequency must be swept slowly.

A new receiving technique, which was proposed by Dr. J. L. Pawsey, is at present under development. In this system, the frequency-sweeping method is replaced by a fixed-frequency system, in which the output of a broad-band

* An account of this survey was presented to a meeting of the American Astronomical Society, held at Boulder, Colorado, in August 1953 (abstract, Kerr and Hindman 1953).

intermediate-frequency channel is analysed by filtering it through a number of parallel narrow-band channels, tuned to different frequencies across the range of interest. The outputs of these channels can all be displayed simultaneously, so that a picture of the whole frequency profile can be obtained in a single integration period. The suppression of zero-level variations and removal of the background are in this case effected by comparing the mean noise level from each narrow-band channel with that in the full receiver bandwidth.

Preliminary tests of the principles involved in this receiver are being carried out with a limited number of channels. The observations described in this paper were made for the most part with a single narrow channel of 40 kc/s bandwidth.

The use of this type of receiver leads to a different method of sweeping over the relevant coordinates. With the usual technique, the aerial follows a fixed point in the sky, while the receiver is swept slowly over the appropriate frequency range, thus tracing out the line profile at that point. The distribution over the sky is then obtained by fitting together a series of these profiles.

In the new system, the aerial remains in a fixed position, a continuous record being obtained at a constant frequency as the sky moves past the aerial beam. In the current observations, line profiles are derived by fitting together a series of these constant-frequency records. The later system will give a simultaneous output at a large number of separate frequencies, yielding the line profile directly as a function of position in the sky.

The parabolic aerial which is being used in this work is shown in Plate I. It has a diameter of 36 ft, and is mounted on an east-west axis as a transit instrument. The beam width between half-power points is 1.0° . This may be taken as the angular resolution of the equipment.

The resolution in frequency is also an important parameter. This depends on the bandwidth of the narrow channel, 40 kc/s, corresponding to a velocity spread of 9 km/s.

(b) *Method of Measurement*

An individual observation involves the measurement of a brightness temperature for a given region of the sky and a given frequency (or velocity). The methods used in measuring the brightness temperature, the direction of the aerial beam, and the frequency will be discussed in turn.

(i) *Brightness Temperature*.—The noise fluctuations appearing in the receiver output have been used as a means of calibrating the output scale in terms of aerial temperature. For a receiver of the type described above, the temperature change, ΔT , which corresponds to the r.m.s. noise fluctuation, is related to the noise factor by the expression

$$\Delta T = \frac{2^{\frac{1}{2}}(N-1)T_0}{(\tau B_c)^{\frac{1}{2}}},$$

where N is the noise factor for the continuous spectrum, T_0 the ambient temperature, τ the output time constant, and B_c the bandwidth of a single narrow channel. When the total receiver bandwidth is much greater than B_c , the

noise arising in the wide comparison band can be neglected in comparison with that due to the narrow band.

The derivation of the above equation, and its relation to the corresponding equations for the systems used by Dicke (1946) and by Ewen and Purcell (1951) will be given in a later paper.

The noise factor of the receiver was measured by means of a gas-discharge noise generator, whose characteristics had been checked against the standards established at a wavelength of 10 cm by the Division of Electrotechnology, C.S.I.R.O.

During the observations described in this paper, the receiver parameters were $N=6$, $\tau=15$ sec, $B_c=40$ kc/s, giving a calculated temperature fluctuation of 2.4°K . Some unwanted slow variations were, however, superposed on the noise fluctuations, reducing the effective receiver sensitivity. For calibration a smaller value of τ was used in order to generate larger noise fluctuations, comparable in amplitude with the observed signals. Use of the larger calibrating signal also minimized errors arising from receiver instability. During the observations, a relative zero point was established at frequent intervals throughout each run by moving to the south celestial pole. At the frequencies used for observation of the Magellanic Clouds, the radiation in the polar region is negligibly small.

Since the mixer and 30 Mc/s preamplifier were mounted at the aerial feed point, it was not convenient to measure the receiver noise factor at frequent intervals. The relative sensitivity of the receiver was therefore checked after each day's run by observing the intensities at a number of points near the galactic plane.

In deriving temperatures by the procedure outlined above, there are two types of uncertainty. The percentage error of the scale calibration is about ± 20 per cent. Relative values are, however, mainly affected by the uncertainty of the zero level. In this case the estimated probable error is $\pm 3^\circ\text{K}$.

All values quoted refer to the "aerial temperature", and the small correction to allow for losses in the aerial and reception from directions other than that of the main beam has not yet been determined.

Temperatures measured in the differential output of the receiver are subject to a further correction. The wide channel in the present equipment has a bandwidth of only 1.7 Mc/s, so that the presence of line radiation will produce a small increase in the wide-band output in addition to the increase in the narrow-band output. This effect must be allowed for in deriving the "true" profile.

(ii) *Direction of Aerial Beam*.—The declination scale, which was attached to the mechanical axis of the aerial, was set by making radio observations of the Sun. If the aerial structure maintains the same shape in all positions, such a single-point calibration will suffice for all declinations. It was found however that structural distortions in the framework and in the feed support resulted in directional errors of up to 1° . Such errors are of course serious for a 1° beam. Two types of measurement were made to determine the distortion. Firstly, the transit of stars at various declinations was observed visually, using a small

telescope rigidly attached to a point on the aerial framework. Secondly, direct measurements were made of the relative movements occurring during rotation of the aerial, by sighting with a telescope from one part of the moving structure to another.

The indicated declinations were corrected according to the results of these measurements. There may still, however, be small systematic errors in both declination and Right Ascension.

(iii) *Frequency*.—The receiver is of superheterodyne type with a double frequency change. A simplified block diagram is shown in Figure 1. The first intermediate-frequency channel is at 30 Mc/s, the second at 7 Mc/s. The overall bandwidth is 1.7 Mc/s, and a portion of this, 40 kc/s wide, is selected

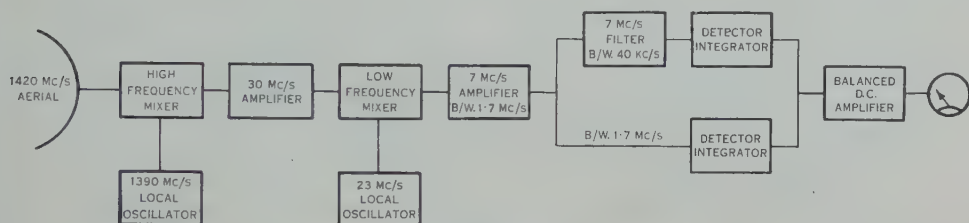


Fig. 1.—Block diagram of receiver.

by the narrow-band filter. Although normally fixed, the operating frequency of the receiver can be varied over a range of several megacycles per second by tuning the first local oscillator, whose frequency is approximately 1390 Mc/s. This frequency is derived from a temperature-stabilized self-excited oscillator operating at 5.7 Mc/s.

Measurement of the receiver frequency involves measurement of its three components, the frequencies of the two local oscillators (1390 Mc/s and 23 Mc/s) and the narrow-band filter (7 Mc/s). With an overall passband of 40 kc/s, an accuracy of several kilocycles per second is required in each case. Thus only the first of the three measurements offers any difficulty. The method adopted makes use of a crystal-controlled oscillator (occasionally checked against WWV) as a local reference standard, to set the self-excited oscillator controlling the first injection frequency.

The latter frequency is determined by a measurement of the audible beat note between these two oscillators. A convenient arrangement is to adjust the various frequencies involved to give zero beat when the receiver frequency corresponds to zero radial velocity (1420.405 Mc/s). Then, for other settings of the receiver, the beat note is a direct measure of the radial velocity at which observations are being made.

III. OBSERVATIONAL RESULTS

(a) *Method of Scanning*

In each run, the receiver was kept on a constant frequency and observations made as the sky moved past the aerial beam. Since the motion is quite slow near the pole, relative to the beam width, it was possible to vary the declination

in steps throughout each run. The result of each run was therefore a map showing the brightness distribution over one or both Clouds at a single frequency, that is, the distribution of the radiation from that part of the gas which is moving with a particular radial velocity. Figure 2 shows in contour form some samples of these maps for the Small Cloud.* The series illustrates the striking change in the appearance of the Cloud for different radial velocities. The values of radial velocity have been referred to the Sun.

In these observations, the galactic foreground makes no contribution in the frequency range of interest; the great difference in radial velocity completely separates the radiation from the Clouds and the Galaxy.

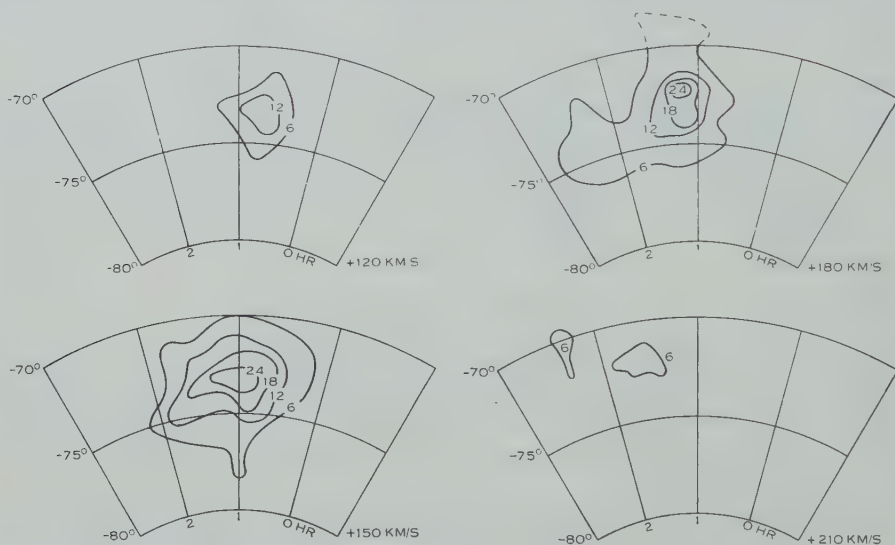


Fig. 2.—Distribution of brightness temperature over the Small Cloud at four radial velocities. (Unit = 1 °K.)

(b) Line Profiles

From a set of maps of this type, line profiles were drawn for 250 points of a lattice, spaced 1° in declination and 10 min in Right Ascension. A number of sample profiles are shown in Figure 3,* indicating the main features of the variation over the Small Cloud.

It can be seen that the brightness temperatures are in all cases low. The highest value observed is about 30 °K, a quarter of the peak brightness temperature at the galactic anticentre.

The group of profiles illustrates the progressive change of velocity across the Clouds, and also shows the different shapes which were obtained. The profiles are fairly smooth and symmetrical over a small region in the immediate centre of the Small Cloud, but elsewhere are generally complex. They are in

* The correction for the contribution of the line to the wide-band output has not been applied in these two figures. The true profiles will be very similar in shape, but the brightness temperatures and linewidths should both be increased by about 15 per cent.

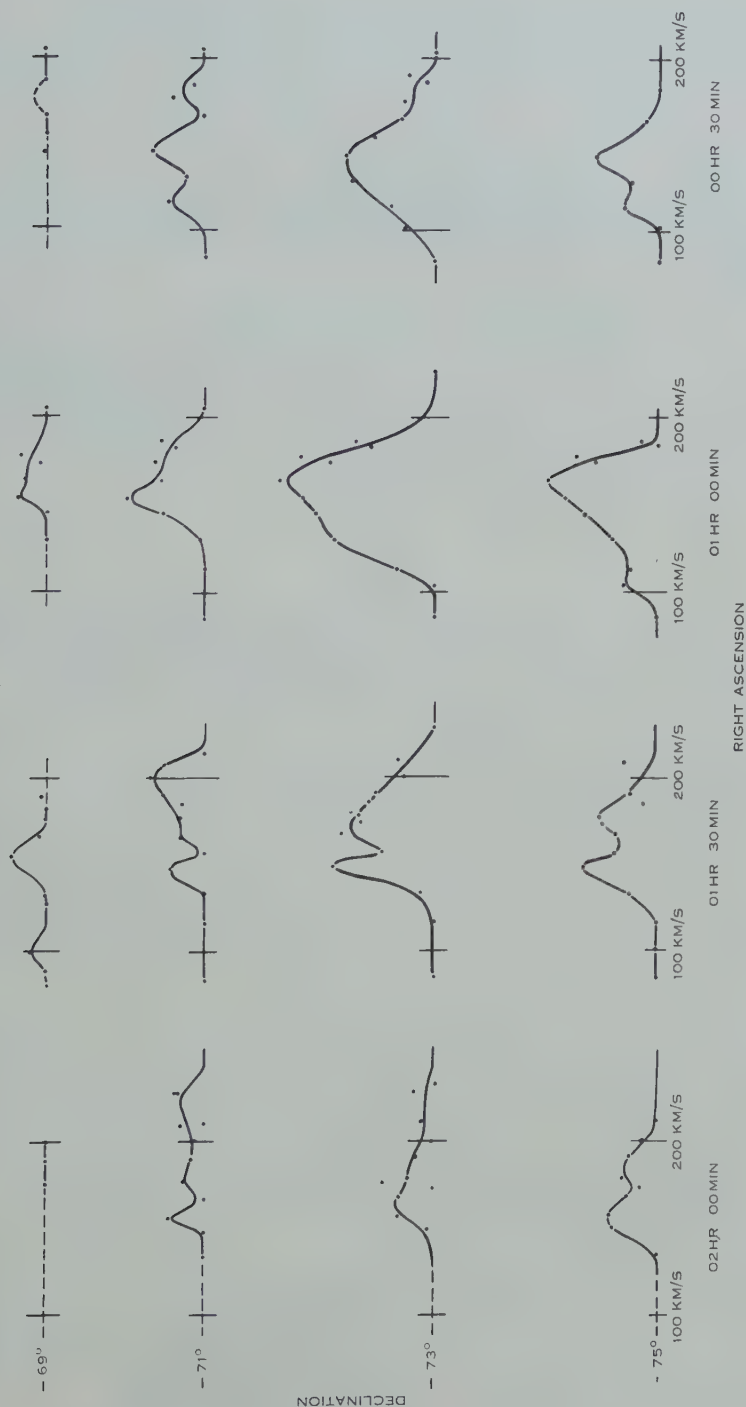


Fig. 3.—Samples of the line profiles obtained for the Small Cloud. The maximum value of the brightness temperature is 30 °K.

general more irregular for the Large Cloud than for the Small Cloud, presumably indicating a greater degree of fine structure in the space distribution.

The linewidths, between half-power points, average about 250 kc/s, corresponding to a velocity spread of 50 km/s. There is no great variation in the linewidth over the region of the Clouds, but the average width for the Small Cloud is a little greater than that for the Large Cloud.

In this paper we have neglected the fine details shown by some of the profiles. The conclusions presented are based on the broader features, as specified by two quantities, "integrated brightness" and "mean radial velocity", defined below.

(c) *Contours of Integrated Brightness*

The total energy received from a particular direction can be measured in terms of the area under the line profile. This quantity, which corresponds to the "integrated intensity" of optical spectroscopy, we call integrated brightness. In conformity with the general system of units used in radio astronomy, the integrated brightness, $B_{\text{int.}}$, is defined here as

$$B_{\text{int.}} = \frac{2k}{\lambda^2} \int T_{\nu} d\nu,$$

where the factor 2 covers the two polarizations of the total radiation, and k is Boltzmann's constant, λ the wavelength, T_{ν} the brightness temperature, and ν the frequency. The m.k.s. unit is the watt(metre)⁻²(steradian)⁻¹. For example, a brightness temperature of 10 °K over a frequency range of 200 kc/s gives $B_{\text{int.}} = 12 \times 10^{-16}$ W m⁻² sterad⁻¹.

A diagram showing the contours of integrated brightness is given in Plate 2, Figure 1, where the contour interval is 7×10^{-16} W m⁻² sterad⁻¹. The values have been increased by 35 per cent. to allow for the contribution of the line to the wide-band output. A dashed line is also included in the figure to indicate the approximate outer limits from which radiation could be detected with the present sensitivity. The limiting factor in these observations was the uncertainty in the position of the zero level.

(d) *Contours of Radial Velocity*

The mean radial velocity for each profile is defined in the sense of the weighted mean, as the velocity for which the ordinate halves the area under the profile. Contours of mean radial velocity, in kilometres per second, are given in Plate 2, Figures 3 and 4. Positive values of radial velocity indicate recession.

The velocities are least accurate where the radiation is very weak. These contours are restricted to the region in which the integrated brightness exceeds 7×10^{-16} W m⁻² sterad⁻¹.

IV. THE MAGELLANIC CLOUDS

Before discussing the radio observations we shall briefly review the results of optical studies of the Clouds.

The Magellanic Clouds are the nearest external galaxies. Subtending angles of several degrees, they are prominent features of the southern sky, but

are invisible from the greater part of the northern hemisphere. The mean of several recent determinations of their distance (Gum and de Vaucouleurs 1953 ; Shapley 1953 ; Thackeray and Wesselink 1953 ; Shapley and Nail 1954) is 46 kpc, in the new cosmic distance scale.* The two Clouds appear to be at practically the same distance from the Sun, to within the observational uncertainty.

The distances of all other members of the local group of galaxies are considerably greater than this. Hence the two Clouds and the Galaxy appear to form a close group of their own. The Clouds have in fact often been referred to as "satellites" of our Galaxy. de Vaucouleurs (1954) has collected evidence suggesting the existence of a connexion between the Large Cloud and the Galaxy. This would be analogous to the cases reported recently by Zwicky (1953), in which filaments of intergalactic matter connect neighbour members in pairs or groups of galaxies.

A photograph of the Clouds is shown in Plate 2, Figure 2. The Large Cloud has a central elongated section, known as the "axis", surrounded by a number of irregularly situated luminous patches. The small Cloud is more concentrated into a single body, but it too shows some irregularities of structure.

To the unaided eye, the diameters of the Large and Small Clouds are approximately 7 and $3\frac{1}{2}^\circ$. Microdensitometer studies and star counts have considerably extended the recognized dimensions of the Clouds. Shapley and Mohr (1932) obtained 12 and 8° respectively for the diameters averaged over all position angles. The outer isophotes of the Large Cloud are reasonably smooth, but those of the Small Cloud show an extension or "wing" towards the Large Cloud. Shapley (1940) showed from star counts that this feature is at least $1\cdot5^\circ$ in width (in declination), and extends 3 or 4° in Right Ascension to a distance of $6\cdot5^\circ$ from the centre of the Small Cloud, that is, one-third of the way to the centre of the Large Cloud. Star counts by McCuskey (1935) confirmed the increased dimensions of the Large Cloud. He obtained a mean diameter of 11° , and this has recently been extended to 20° by de Vaucouleurs (1954).

Holmberg (1952) has made estimates of the masses of the Clouds. His values, reduced to the new distance scale, are as follows: for the Large Cloud, 2×10^9 solar masses as derived from the velocity dispersion of nebulae observed by Wilson (1917), and $2\cdot5 \times 10^9$ solar masses from the estimated mass/luminosity ratio for a type I population and the Shapley-Ames magnitude; for the Small Cloud, by the latter method, 6×10^8 solar masses. These values must, however, all be taken as rather uncertain.

In the standard classification of galaxies, the Magellanic Clouds are the typical examples of a class of irregular galaxies, which comprises about half of the systems classified as "irregular". H. C. Russell (1890) first reported evidence of spiral structure in the Clouds. Shapley (1931, 1950) described similarities in the Large Cloud and the barred spirals, but he considered that the evidence was not yet clear as to whether the Clouds are flattened or spherical.

* As recently revised, following the discussion presented by Baade at the Rome Assembly of the International Astronomical Union, September 1952.

A more detailed study has been made by de Vaucouleurs (1954) from star counts and multiply printed photographs. He has been able to show the existence of an extensive spiral structure in the outer parts of the Large Cloud. From the shape of the intermediate and outer isophotes, he suggested that both Clouds are flattened systems with different inclinations to the line of sight. The Large Cloud appears to have an inclination of 65° , with its major axis in position angles* $160\text{--}340^\circ$, the Small Cloud an inclination of 35° , and major axis in position angles $40\text{--}220^\circ$.

The stellar population types of the Clouds have been the subject of considerable recent discussion. Baade (1950) has given the Large Cloud as an example of a system with a population of pure type I. Thackeray and Wesselink (1953) have shown, however, that it contains some variables which belong to type II. It is therefore not of pure type I, but appears to be predominantly so—its colour is that of a type I system.

Shapley (1950 and personal communication) has pointed out that the Small Cloud approaches in character the irregular dwarf galaxies and it may be evolving towards the spheroidal type of galaxy, that is, it may be more developed towards one of the Population II subdivisions than is the Large Cloud. Gascoigne and Kron (1953), from colour measurements of the Clouds and some of their variables, concluded that the Small Cloud might be predominantly of type II. Gascoigne (personal communication) now considers that further measurements suggest that this view should be substantially modified, and that the Small Cloud appears to differ in important respects from both Population I and Population II. Shapley (1950) suggests that the population of the Small Cloud, at its centre, may be intermediate between Populations I and II.

The dust content of the two Clouds appears to be very different. Photographs show many more dark patches in the Large Cloud than in the Small Cloud. Shapley (1951) has made counts of the numbers of distant galaxies visible through various parts of the Clouds, and finds that the Small Cloud is essentially transparent, whereas the Large Cloud contains a considerable amount of obscuring dust. He points out, however, (personal communication) that, while the Large Cloud undoubtedly contains more dust, no quantitative comparison is possible because of irregularities in the distribution of the faint galaxies themselves. In particular, there is probably some obscuration in a small region at the centre of the Small Cloud. Recent colour measurements of B stars and cepheids by Gascoigne (personal communication) do not support the presence of appreciable quantities of dust in the Small Cloud: they cannot be said, however, definitely to exclude this possibility.

Ionized hydrogen is known to be present in both Clouds. Studies of emission nebulosities have been reported by Henize and Miller (1950) for both Clouds, and by Shapley and Wilson (1925) and Nail, Whitney, and Wade (1953) for the Small Cloud. Hydrogen emission appears to be more prominent in the case of the Large Cloud.

* The position angle is the inclination to the hour circle passing through the centre of the object, measured from north towards east.

In addition, the H and K lines of interstellar calcium in the Large Cloud have recently been observed by Feast (1953) in the light of S Dor, the brightest star in the Cloud. This is the first observation of absorption by interstellar gas in an extragalactic system.

The information available from optical sources about the motions of the Clouds is scanty. So far, radial velocities have been measured for only 17 bright nebulae in the Large Cloud, and one in the small Cloud (Wilson 1917). The values obtained are in the range $+251$ to $+309$ km/s in the Large Cloud, and $+168$ km/s in the Small Cloud. Proper motions have not yet been measurable.

Wilson examined the radial velocities for evidence of rotation of the Large Cloud, but the data were too few for a clear result. Hertzsprung (1920) then suggested that the data could be interpreted more simply as indicating a translational motion of the two Clouds through space. A more recent analysis has been made by Wilson (1944), leading to a suggested space motion of about 550 km/s.

There are still many gaps in present knowledge and the solution of some pressing problems requires more data. Radio and optical observations are to a great extent complementary. It seems likely that the detailed data of both types, which can be expected in the near future, should lead to a much better understanding of the two systems.

V. DISCUSSION OF THE BRIGHTNESS OBSERVATIONS

We can now discuss the current radio observations in the light of the present optical knowledge of the Clouds. Since this is a preliminary survey, the discussion will largely indicate the type of information which radio studies can provide, rather than derive final results.

(a) *The Size of the Clouds*

The most striking aspect of the brightness contour diagram (Plate 2, Fig. 1) is the large size of the Clouds, in comparison with the size of the easily visible regions (see Plate 2, Fig. 2). Also the large amount of radiation from the Small Cloud, relative to that from the Large Cloud, would not have been expected from its smaller optical luminosity and lower dust content.

Features such as the wing of the Small Cloud and the axis of the Large Cloud can be clearly seen in the diagram. Another noticeable characteristic is that the contour pattern is smoother in the case of the Small Cloud. This conforms with the optical evidence for a more regular structure.

(b) *The Degree of Central Concentration*

The radio brightness is seen to decrease less rapidly than the optical brightness in going outwards from the centre, demonstrating that the gas is less concentrated towards the centre than the bright stars. This result is made more evident by comparing radio and optical brightnesses along sections through the Clouds. Such a comparison is shown in Figure 4, which gives sections at constant declination and constant Right Ascension through the optically brightest part

of each Cloud.* The optical curves have been derived from data obtained by van Herk (1930). These observations are not good by modern standards, but they are the only systematic data available so far. A major weakness is that the brightness at points in the Clouds has been referred to the brightness at nearby points, which were at that time considered to be outside the Clouds. Present knowledge of the size of the Clouds indicates that the reference level cannot be taken as a true zero level. The shape of the optical sections in the figure may therefore be considerably in error in the outer parts, but the central regions will not be greatly affected.

To facilitate the comparison of the radio and optical data, the resolution of the latter has been degraded by "smoothing" the data with a directional characteristic equivalent to that given by an aerial with a beam width of 1° between half-power points. Thus the radio and optical results have been made directly comparable. The two brightness scales have been adjusted to make the maxima coincide for the Large Cloud.

The greater spread of the radio distribution can be clearly seen in both sections through the Small Cloud and in the section at constant Right Ascension through the Large Cloud. Thus the hydrogen appears in each case as a large envelope surrounding a core of bright stars. However, the radio and optical distributions are more similar for the section at constant declination through the Large Cloud, passing along the axis. The elongated axis, which is a feature of the optical picture, appears to be less marked in the radio observations. This suggests that the gas is disposed in a broad distribution, with an approach to circular symmetry, on which the axis of bright stars is superposed.

Detailed studies of brightness distributions of this type are important in relation to the dynamics and the gravitational equilibrium of the Clouds. When more radio data are available it will be possible to study the Clouds in the way which has been applied to globular clusters (for example, Camm 1952) and ellipsoidal galaxies (de Vaucouleurs 1953). These studies are expected to show the extent to which the Clouds have reached a state of equilibrium under their own gravitation and internal motions.

Differences in distribution, similar to that now reported for the hydrogen and the bright stars, are already known in other galactic systems for stars of different sizes. The internal structure of galaxies and clusters was at first deduced solely from studies of the light distribution across their surfaces. As in the case of the Magellanic Clouds, refinements of technique have revealed more and more of the outlying regions, and thus increased the apparent sizes of the systems. In some galaxies the distribution of rotational velocity has now been observed, leading to the mass distribution. In each case the mass and light distribution are found to be quite different (Oort 1940; Wyse and Mayall 1942), with the ratio of mass to luminosity increasing steadily on going out from the nucleus.

* This is not the same as the centroid of the hydrogen distribution, which is at R.A. 05 hr 35 min, Dec. -68.5° for the Large Cloud, and R.A. 01 hr 20 min, Dec. -72.5° for the Small Cloud.

This appears to indicate that most of the mass is contained in dwarf stars, widely distributed around the nucleus. Since the luminosity of a star increases

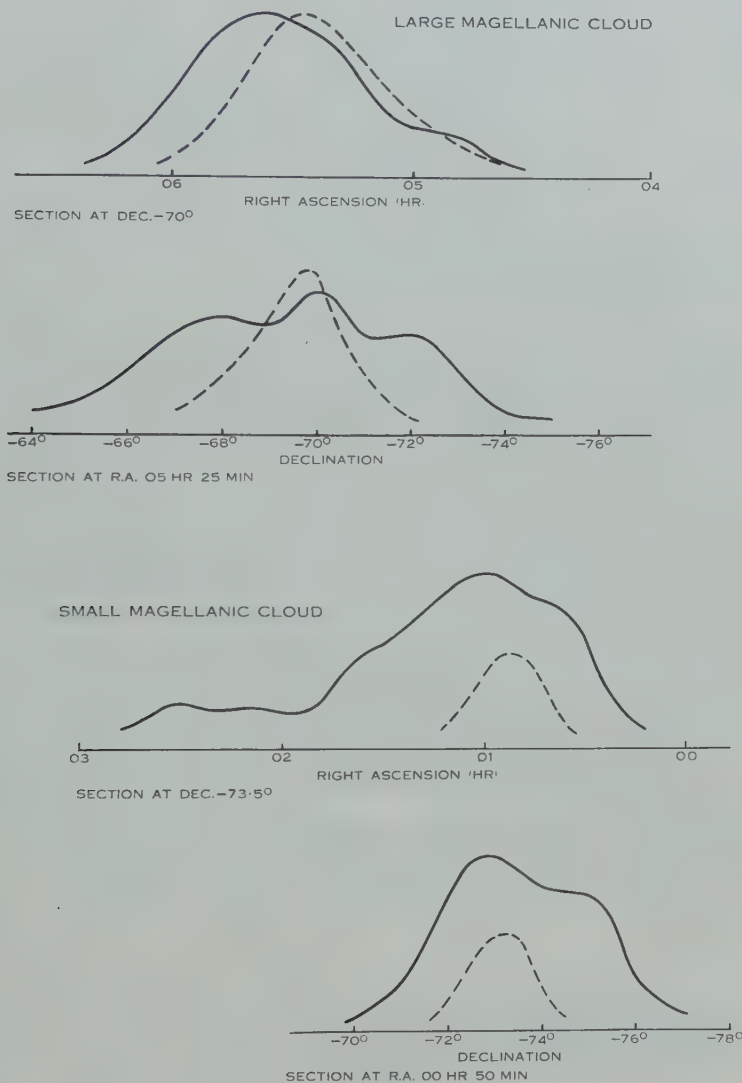


Fig. 4.—Comparison of radio and optical brightness distributions along sections through the Clouds. The brightness scales are the same for the two Clouds, and the radio and optical scales have been adjusted to equalize the peak values for the Large Cloud.

———— Radio. - - - - Optical.

much more rapidly than its mass ($L \propto M^{3.8}$), optical observations tend to pick out the brighter stars, in spite of their smaller number. Evidently these are more concentrated towards the centre than the dwarf stars. Dynamical argu-

ments are used to account for these observed differences in the distribution of stars of different masses.

In the case of the Clouds, then, the main mass is likely to be in dwarf stars, more spread out than the bright stars which mainly contribute to the observed luminosity. The hydrogen distribution may therefore approximate to the mass distribution, but later work should enable a detailed comparison to be made.

(c) *Quantity of Neutral Hydrogen*

Unless the source region is opaque, the observed integrated brightness (the area under the profile) will increase with an increase in the number of hydrogen atoms in the line of sight. When the optical depth is small, the relationship is nearly linear, and is also independent of the excitation temperature of the source. (This latter result follows from the fact that the absorption coefficient is inversely proportional to the excitation temperature.) In these circumstances, the quantity of hydrogen in the line of sight can be derived directly from the observational data and an integration over the brightness contours can lead to the total mass of neutral hydrogen in each Cloud.

This method cannot be used, however, when the number of atoms—and so the optical depth—is very great, because then the presence of additional atoms does not increase the brightness temperature. The line is then said to be “saturated”, the increased emission due to the additional atoms being balanced by increased self-absorption. In the case where the radiating gas is at a uniform temperature, the resulting profile would be “flat-topped”, though this may not be so in the general case.

The shapes of the observed profiles and the low values of the brightness temperatures provide some evidence that the Clouds are optically thin. Therefore the values for the quantity of hydrogen which are derived below from the integrated brightness are probably the correct figures and are certainly lower limits.

The required expression for the number of atoms in the line of sight can be derived from the equation for the absorption coefficient (after Wild 1952),

$$\kappa = 2.6 \times 10^{-15} \cdot \frac{n}{\theta} \cdot f(\nu) \text{ cm}^{-1},$$

where n is the number of ground-state hydrogen atoms per cubic centimetre, θ is the excitation temperature in degrees Kelvin, and $f(\nu)$ is the “line-shape” function, normalized so that

$$\int_0^\infty f(\nu) d\nu = 1.$$

The optical depth is then

$$\tau = \int_0^\infty \kappa ds = 2.6 \times 10^{-15} \int_0^\infty \frac{n}{\theta} f(\nu) ds,$$

and, if θ and $f(\nu)$ are assumed to be constant throughout the depth s of the gas in the line of sight,

$$\tau = 2.6 \times 10^{-15} \frac{f(\nu)}{\theta} N,$$

where N is the total number of atoms in the line of sight in a cylinder of 1 cm^2 section.

The brightness temperature of the line is then given by

$$T_v = \theta(1 - e^{-\tau}) \\ \simeq \theta\tau, \quad \text{when } \tau \ll 1.$$

Integrating the brightness temperature over the full range of ν , we have

$$N = \frac{\int_0^\infty T_v d\nu}{2.6 \times 10^{-15}} = \frac{\lambda^2 B_{\text{int.}}}{5.2 \times 10^{-15} k} \\ \text{(from the definition of integrated brightness)} \\ = 6.2 \times 10^{35} B_{\text{int.}}$$

This is the number of atoms in a projected area of 1 cm^2 . An area of 1 deg^2 on the celestial sphere at a distance of 46 kpc is equal to $6.2 \times 10^{42} \text{ cm}^2$. Thus the number of atoms corresponding to an integrated brightness $B_{\text{int.}}$ extending over an area $\delta A \text{ deg}^2$ is

$$\delta N = 3.8 \times 10^{73} B_{\text{int.}} \delta A.$$

The total number of neutral hydrogen atoms in each Cloud can now be found by integrating the product ($B_{\text{int.}} \delta A$) over the contour diagram for the Cloud. The values obtained are 7×10^{65} and 5×10^{65} for the Large and Small Cloud respectively. The corresponding masses of hydrogen are 6×10^8 and 4×10^8 units of solar mass.

These results may be compared with Holmberg's estimates of the masses of the Clouds (see Section IV), 2×10^9 and 2.5×10^9 solar masses for the Large Cloud by two methods, and 6×10^8 solar masses for the Small Cloud.

In addition to the mass of neutral hydrogen, better values for the total mass of each Cloud should be obtainable from hydrogen line observations (see Section VI). The present data lead to a tentative value of 20 per cent. for the proportion of hydrogen, by mass, in both Clouds. This may be compared with a recent estimate of 15 per cent. for this ratio in the solar neighbourhood (Oort 1952).

More reliable data will make possible a detailed comparison between the distribution of the gas through the Cloud and the distribution of (total) mass. The latter would be derived from the velocity distribution.

An important result from this survey is that the amount of gas in the two Clouds is about the same, whereas the Large Cloud is believed to contain much more dust than the Small Cloud. Thus the ratio of gas to dust, which has often been held to be constant everywhere, is very different in the two systems. Better optical data are required to put this result on a quantitative basis, but qualitatively it appears to be well established.

The mean gas density can also be derived. In the central regions of the two Clouds the number of atoms in a cylinder of 1 cm^2 section in the line of

sight is about 3×10^{21} . The mean density of neutral hydrogen is then of the order of $10^3/d$ atoms/cm³, where d is the "depth" of either Cloud in parsecs. The mean diameter of the hydrogen distribution for the Large Cloud, in a plane perpendicular to the line of sight, is 5 kpc between half-intensity points. Thus, if the Cloud has an approximately spherical shape, the mean density in the central region would be about 0.2 atoms/cm³. On the other hand, if we take the conventional value of 1 atom/cm³ derived from observations in the solar neighbourhood, we obtain a value of 1 kpc for the line-of-sight depth. The corresponding values for the Small Cloud are very similar.

VI. DISCUSSION OF THE VELOCITY OBSERVATIONS.

The observations on velocities will now be considered. The radio observations have added enormously to the available information on velocities in the Clouds. Optical measurements to date only give radial velocities for 17 bright nebulae in the Large Cloud, and one in the Small Cloud. The present radio data give velocities for more than 200 independent points across the Clouds, and for each point a full velocity profile can be obtained. The spread of velocity in a profile (about 50 km/s between half-intensity points) can be due to random, or peculiar, motions in the Clouds, and also to systematic variation of velocity with depth in the Cloud, along the line of sight.

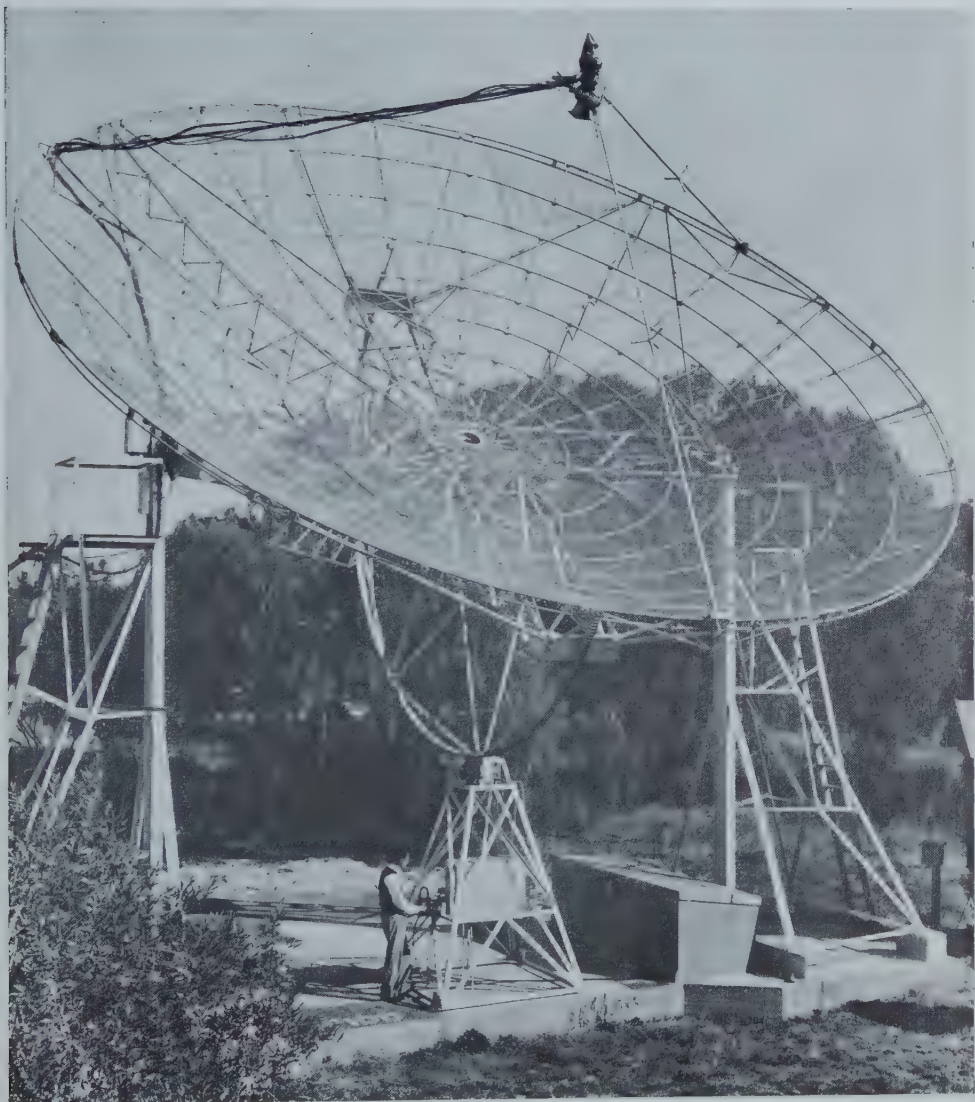
The present survey was not sufficiently detailed to give the fine structure of the velocity profiles. Systematic motions can, however, be studied through the mean radial velocity, defined in Section III (*d*). The significance of this quantity is that, provided the optical depth is small, it gives the motion of the centre of mass of the material in the line of sight. In the particular case of a flattened system it corresponds, on the average, to the mean motion in the equatorial plane.

The contour diagram of mean radial velocity, presented in Plate 2, Figure 3, shows a systematic pattern, with velocity increasing from right to left. The greater part of these velocities, and of their variation across the Clouds, arise from the Sun's rotation about the centre of the Galaxy. The next step is therefore to remove the known motions.

Plate 2, Figure 4, shows the detailed mean radial velocity data, after removal of the galactic rotation velocity (taken as 270 km/s towards $l=57^\circ$, $b=0^\circ$) and the solar motion relative to the local centre of mass (20 km/s towards R.A. 270° , Dec. 30.0° (1900)). The residual velocities are found to cover a range between -45 and $+60$ km/s. Analysis of the distribution of these residual velocities indicates that each Cloud is rotating. The orientation of the major axis agrees with that derived from the shape of the optical isophotes (see Section IV). The rotational parameters lead to values for the total masses of the Clouds. Thus we can obtain from hydrogen line observations both the mass of neutral hydrogen and the total mass.

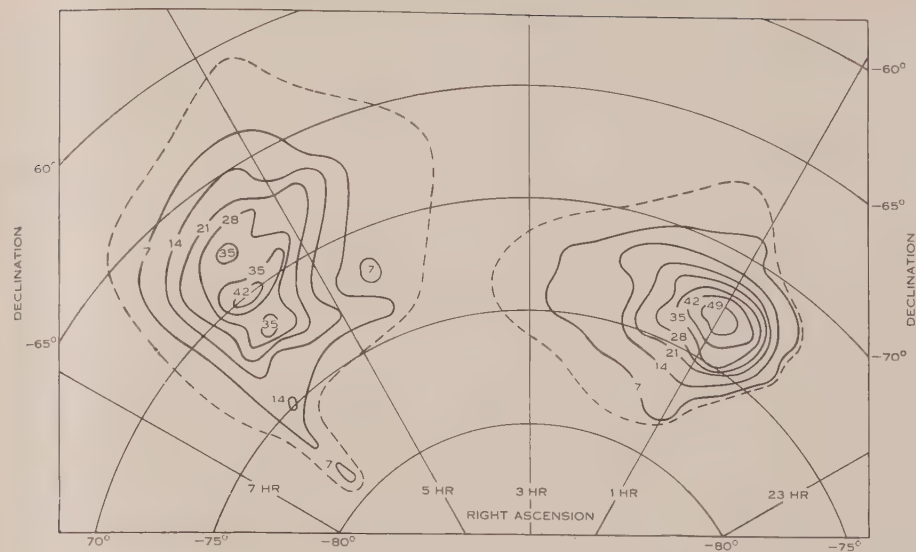
Weighted mean radial velocities have been calculated, representing the velocity of the centre of mass of the neutral hydrogen in each Cloud. The

21 CM LINE FROM MAGELLANIC CLOUDS



The 36-ft. paraboloid at Potts Hill, near Sydney.
The aerial is mounted as a transit instrument.

21 CM LINE FROM MAGELLANIC CLOUDS



values obtained are +280 and +160 km/s for the Large and Small Cloud respectively, relative to the Sun, and +37 and -16 km/s after removal of the Sun's motions.

A detailed discussion of the velocity data in Plate 2, Figure 4, will be given in a subsequent paper.

VII. CONCLUSIONS

The 21 cm line has been received from large areas extending well beyond the easily visible regions of each Cloud, and the hydrogen of the Clouds is much less concentrated towards the nucleus than are the brightest stars.

Despite the very different quantities of dust in the two Clouds, each system contains about the same total amount of gas. Thus the ratio of gas to dust is very different in the two Clouds.

If the Clouds are taken to be optically thin for this radiation, the neutral hydrogen content is found to be 6×10^8 and 4×10^8 solar masses for the Large and Small Cloud respectively.

The results have illustrated the great power of 21 cm line studies. As well as giving the total quantity of gas and its distribution, the method can very quickly supply information about radial velocities within the Clouds in more detail than was available from previous optical studies.

VIII. ACKNOWLEDGMENTS

The authors wish to thank Dr. G. de Vaucouleurs, Research Fellow of the Australian National University, for valuable discussions; Dr. J. L. Pawsey for his stimulating influence as leader of the radio astronomy group at the Radio-physics Laboratory; and Mr. C. F. Fryar for assistance with the equipment and the observations.

IX. REFERENCES

- BAADE, W. (1950).—*Publ. Obs. Univ. Mich.* **10**: 7-17.
 CAMM, G. L. (1952).—*Mon. Not. R. Astr. Soc.* **112**: 155-76.
 CHRISTIANSEN, W. N., and HINDMAN, J. V. (1952).—*Aust. J. Sci. Res. A* **5**: 437-55.
 DICKE, R. H. (1946).—*Rev. Sci. Instrum.* **17**: 268-75.
 EWEN, H. I., and PURCELL, E. M. (1951).—*Nature* **168**: 356.
 FEAST, M. W. (1953).—*Mon. Not. Astr. Soc. S. Afr.* **12**: 63-4.
 GASCOIGNE, S. C. B., and KRON, G. E. (1953).—*Publ. Astr. Soc. Pacif.* **65**: 32-6.
 GUM, C. S., and DE VAUCOULEURS, G. (1953).—*Observatory* **73**: 152-5.
 HENIZE, K. G., and MILLER, F. D. (1950).—*Publ. Obs. Univ. Mich.* **10**: 75-8.
 VAN HERK, G. (1930).—*B.A.N.* **6**: 61-4.
 HERTZSPRUNG, E. (1920).—*Mon. Not. R. Astr. Soc.* **80**: 782-4.
 HOLMBERG, E. (1952).—*Medd. Lunds Astr. Obs. Series I*, No. 180.
 VAN DE HULST, H. C. (1945).—*Ned. Tijdschr. Natuurk.* **11**: 201.
 VAN DE HULST, H. C. (1953).—*Observatory* **73**: 129-39.
 KERR, F. J., and HINDMAN, J. V. (1953).—*Astr. J.* **58**: 218-9.
 MCCUSKEY, S. W. (1935).—*Circ. Harv. Astr. Obs.* No. 401.
 NAIL, VIRGINIA MCK., WHITNEY, C. A., and WADE, C. M. (1953).—*Proc. Nat. Acad. Sci., Wash.* **39**: 1168-76.
 OORT, J. H. (1940).—*Astrophys. J.* **91**: 273-306.
 OORT, J. H. (1952).—*Astrophys. J.* **116**: 233-50.
 RUSSELL, H. C. (1890).—*Mon. Not. R. Astr. Soc.* **51**: 39-43, 96-7.

- SHAPLEY, H. (1931).—Bull. Harv. Astr. Obs. No. 881, pp. 1-4.
SHAPLEY, H. (1940).—Bull. Harv. Astr. Obs. No. 914, pp. 8-9.
SHAPLEY, H. (1950).—*Publ. Obs. Univ. Mich.* **10** : 79-84.
SHAPLEY, H. (1951).—*Proc. Nat. Acad. Sci., Wash.*, **37** : 133-45.
SHAPLEY, H. (1953).—*Proc. Nat. Acad. Sci., Wash.* **39** : 349-57.
SHAPLEY, H., and MOHR, JENKA (1932).—Bull. Harv. Astr. Obs. No. 889, pp. 13-15.
SHAPLEY, H., and NAIL, VIRGINIA MCK. (1954).—*Proc. Nat. Acad. Sci., Wash.* **40** : 1-5.
SHAPLEY, H., and WILSON, HARVIA H. (1925).—Circ. Harv. Astr. Obs. No. 276.
THACKERAY, A. D., and WESSELINK, A. J. (1953).—*Nature* **171** : 693.
DE VAUCOULEURS, G. (1953).—*Mon. Not. R. Astr. Soc.* **113** : 134-61.
DE VAUCOULEURS, G. (1954).—*Observatory* **74** : 23-31.
WILD, J. P. (1952).—*Astrophys. J.* **115** : 206-21.
WILSON, R. E. (1917).—*Publ. Lick Obs.* **13** : 187-90.
WILSON, R. E. (1944).—*Publ. Astr. Soc. Pacif.* **56** : 102-6.
WYSE, A. B., and MAYALL, N. U. (1942).—*Astrophys. J.* **95** : 24-47.
ZWICKY, F. (1953).—*Phys. Today* **6** : 7-11.

THE BAROMETER COEFFICIENT AND AIR MASS EFFECTS ON COSMIC RAYS AT MACQUARIE ISLAND

By R. M. JACKLYN*

[*Manuscript received November 16, 1953*]

Summary

The changes in cosmic ray intensity associated with the passage of weather fronts over the observing station have been investigated, using data from Macquarie Island. The effects can be explained in terms of the different way in which the height of the production layer for mesons varies with surface pressure for warm moist and cold dry air masses. The barometer coefficients found for these air masses are respectively -0.120 ± 0.058 per cent. per mb and -0.220 ± 0.041 per cent. per mb for the penetrating component. These results indicate that the frequently observed fluctuations in short-term barometer coefficients may be traceable to changes in air mass types.

I. INTRODUCTION

During the past 10–15 years several workers (Loughridge and Gast 1940 ; Nishina *et al.* 1940*a*, 1940*b* ; Trumpy 1949 ; Lindholm 1950) have commented on the marked changes which cosmic ray intensities undergo with the passage of fronts over the recording station. This effect was studied at the Australian National Antarctic Research Expedition (A.N.A.R.E.) station at Macquarie Island (lat. 54 °S., long. 159 °E.) during 1951.

The weather at the station was typically overcast, with a yearly average humidity of 88 per cent. Daily and seasonal surface temperature fluctuations were comparatively small, but large variations of surface pressure occurred, often accompanied by a change of air mass. (The phrase “air mass”, for the purpose of this study, refers to a characteristic combination of surface and upper air temperatures, humidity, and surface wind direction, which usually persists over the station for several days at a time, and with the passage of a surface of discontinuity, the front, changes to another typical set of the quantities.)

A change of air mass was generally associated with a cold or an occluded cold front. The origin and orientation of fronts and the tracks of major depressions in the Southern Ocean are still matters of conjecture, partly because there are very few observing stations in the area. Gibbs, Gotley, and Martin (1952) have assumed that the major source region of fronts is in the close neighbourhood of the Antarctic Continent. Briefly, they appear to develop as follows. Outbursts of cold dry air move northwards behind Antarctic fronts until they become the polar fronts of middle latitudes. Wave developments in these fronts result in the growth of major depressions. One of the two main tracks followed by

* Member of the Australian National Antarctic Research Expedition at Macquarie Island, 1951–52 ; present address : Physics Department, University of Tasmania, Hobart.

these depressions, and their associated fronts, is in a south-east direction from the south-west coast of Australia to the vicinity of Macquarie Island, reaching maximum development at 60 °S. As they approach the Antarctic coastline and gradually dissolve, the depressions cause fresh outbreaks of Antarctic air to occur, with new Antarctic fronts as their forward boundaries. So it seems likely that cold fronts at Macquarie Island are either well-developed polar fronts associated with depressions tracking from the north-west, or else Antarctic fronts associated with fresh air masses from the Continent.

II. EQUIPMENT AND METHOD OF ANALYSIS

The cosmic ray records used for this work were obtained from two independent units, identical in geometry with those used by Caro, Law, and Rathgeber (1948). But the recording circuits differed, and gave a resolving time for coincidence of 2 μ sec.

A wide angle vertical telescope comprised two closely spaced trays of Geiger counters, each having an effective sensitive area of 400 cm². Double coincidences from this unit measured the wide angle total intensity (referred to as count 3) at a rate of about 20,000 counts/hr. The other unit, a narrow angle vertical array comprised four trays of the same type as above, with a spacing of 15 cm between adjacent trays. Ten cm thickness of lead was placed over the bottom tray so that triple coincidences from the lower trays measured the narrow angle penetrating component (referred to as count 2) at a rate of about 3400 counts/hr. The top three trays were used in triple coincidence to measure the narrow angle total intensity (referred to as count 1) at a rate of approximately 4800 counts/hr. Since counts 1 and 2 were derived from arrays with identical geometry, an approximate measure of the non-penetrating radiation could be obtained by taking the difference between them, but the derived counting rate was found to be too low to give results of sufficient statistical accuracy for this work.

Count 3 (wide angle total intensity), count 1 (narrow angle total intensity), and count 2 (penetrating component) were each analysed for the average effects produced by the passage of fronts, following a procedure used by Loughridge and Gast (1940). For an individual front, the counting rates were collected into 6-hr groups, referred to the time of passage over the station. Thus the 6-hr group immediately preceding the change of air mass was centred about the time 3 hr before the passage of the front. The counting rates were corrected to a standard pressure of 940 mb using the appropriate bi-monthly barometer coefficient, derived from the daily mean rates and surface pressures. To minimize the effects of fluctuations in the general level of intensity between one front and another, the corrected rates were normalized to the second 6-hr group preceding the front. Thus only the mean differences were plotted between this and other groups.

Counts 1 and 2 were analysed for six well-defined cold fronts and five double cold fronts (for each of which the frontal changes occurred in two distinct stages several hours apart) recorded between June and November 1951. These have been treated as a single group of 11 cold type fronts, to obtain a better estimate

of the post-front changes. Over the same period, results from count 3 were available for 14 cold type fronts.

Since the same apparatus was in operation at Macquarie Island during 1950, the records for that year were also analysed for the effect of fronts, and in this case it was decided to consider only the occasions for which both independent units were operating, so that the records for all counts were simultaneous. This restricted the results to 12 marked cold fronts. No double fronts were included in this group.

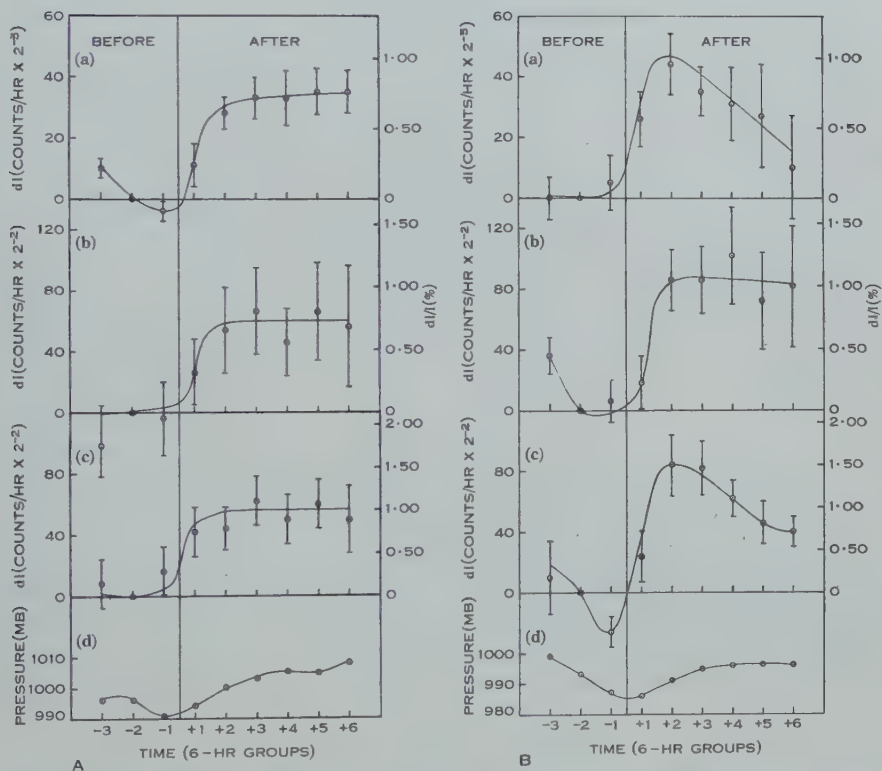


Fig. 1A.—The average effect of the passage of cold fronts during 1950 on (a) the wide angle total intensity (count 3), (b) the narrow angle total intensity (count 1), (c) the narrow angle penetrating component (count 2), together with (d) the mean surface pressure variations. The error tails shown are the 95 per cent. fiducial limits.

Fig. 1B.—The average effect of the passage of cold type fronts during 1951. (a)–(d) as for Figure 1A.

III. THE FRONT EFFECTS

The average effects on the corrected counting rates for each of the years 1950 and 1951 are shown graphically in Figures 1A and 1B, together with the mean surface pressure variations. The post-front increases in the level of corrected scaled counts for each type of telescope as well as the percentage increases are set out in Table 1. It can be seen that the increases are consistent for both years, and they are in qualitative agreement with the ionization chamber

measurements made by Loughridge and Gast. The increases in corrected scaled counts for the penetrating radiation and total radiation are the same for the two telescopes with the same geometry. The smaller percentage increase in total radiation is practically the same for both the narrow angle and the wide angle telescopes. Thus it appears that the post-front effects are almost entirely due to variations in the penetrating radiation.

TABLE I

POST-FRONT INCREASES IN COSMIC RAY INTENSITIES CORRECTED FOR PRESSURE USING A TOTAL BAROMETER COEFFICIENT

	Year	Count 2	Count 1	Count 3
Number of fronts	1950	12	12	12
	1951	11	11	14
Total increase (scaled counts) ..	1950	50-60	50-60	
	1951	80-90	80-90	
Percentage increase	1950	0.93	0.75	0.75
	1951	1.5	1.0	1.0

IV. THE MECHANISM OF THE FRONTAL EFFECTS AND THE VARIABILITY OF THE BAROMETER COEFFICIENT

As Duperier (1949) has pointed out, cosmic ray intensity may be regarded as a function of several meteorological variables. He proposed that the following three were predominant: the surface pressure B , the height H of the pressure level where the majority of mesons are produced, and the mean temperature T in the neighbourhood of this level. Thus a variation of the intensity I is given by

$$\delta I = \mu \delta B + \mu' \delta H + \alpha \delta T.$$

On the other hand, if cosmic ray intensity is regarded as a function of surface pressure only, and

$$\delta I = \beta \delta B,$$

where β is the total barometric regression coefficient, then

$$\beta = \mu + \mu' \frac{\delta H}{\delta B} + \alpha \frac{\delta T}{\delta B} \dots \dots \dots (1)$$

Evidently then, the pressure coefficient varies with the upper air meteorological conditions, and may undergo marked and consistent changes when one type of air mass replaces another.

With this in mind, the total pressure coefficients for the pre-front moist warm air and the post-front polar air for the months June to November were

obtained from the 1951 records. (Unfortunately it was not possible to extract this kind of data for the 1950 results.) The coefficients for counts 1 and 2 are set out in Table 2 (the errors given being the 95 per cent. fiducial limits) together with the bi-monthly coefficients used for the analysis described above. In the passage from a temperate to a polar air mass there is a significant increase in the pressure coefficient for the penetrating radiation, and that for the total intensity shows a similar trend. Notably, all the bi-monthly coefficients have values lying between the two air mass values.

TABLE 2
BAROMETER COEFFICIENTS 1951

	Penetrating Component (counts/mb)	Narrow Angle Total Intensity (counts/mb)
Temperate air mass . . .	-1.018 ± 0.49 ($-0.120 \pm 0.058\%/mb$)	-1.093 ± 0.41 ($-0.184 \pm 0.069\%/mb$)
Polar air mass	-1.877 ± 0.35 ($-0.220 \pm 0.041\%/mb$)	-1.563 ± 0.23 ($-0.260 \pm 0.037\%/mb$)
Bi-monthly means		
May-June	-1.591	-1.44
July-August	-1.426	-1.44
September-October	-1.231	-1.44

If the counting rates are corrected using the appropriate coefficients for the air masses (Fig. 2), it is obvious that the change in corrected rates on passing through a front is arbitrary, depending on the choice of a standard pressure. If, as in the case of count 2, the standard pressure is chosen corresponding to the point of intersection of the regression lines (1005 mb) there should be no change in corrected rates on passing from the temperate to the polar air mass. But it is clear that if a single average pressure coefficient is used, as in the data summarized in Figure 1, its value lying somewhere between the values for the polar and temperate air mass coefficients, the rates corrected to a standard pressure must increase after the passage of a cold front. This accounts for the fact that the pattern of variation of penetrating and total radiation, corrected using the same barometer coefficient before and after cold type fronts, is consistent for both years on Macquarie Island and is in qualitative agreement with the observations made by Loughridge and Gast (1940). This being so, referring to equation (1), the ratios $\delta H/\delta B$ and $\delta T/\delta B$ must vary consistently both in different years and in different places on passing from warm moist to cold dry air.

Now, Trumpy (1949) has accounted for variation in meson counting rate due to the passage of fronts by assuming that the surface pressure B and the height of the 100-mb level H are the predominantly effective variables, so that

$$\delta I = \mu \delta B + \mu' \delta H.$$

The corresponding expression for the barometer coefficient β is

$$\beta = \mu + \mu' \frac{\delta H}{\delta B}.$$

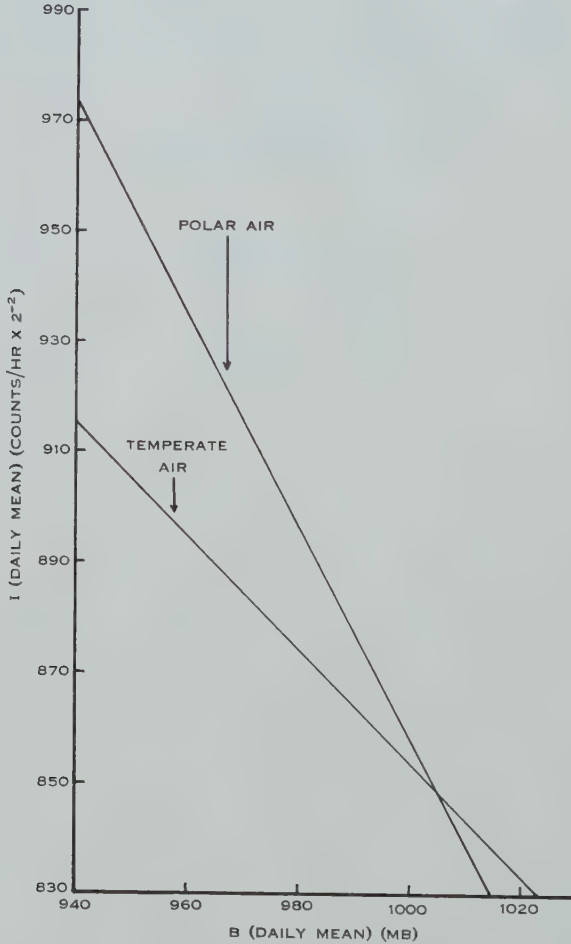


Fig. 2.—The line of regression of penetrating radiation (count 2) on surface pressure under different air mass conditions.

Using the suffixes P and T to denote polar air mass and temperate air mass conditions respectively, we have

$$\beta_P = \mu + \mu' \left(\frac{\delta H}{\delta B} \right)_P, \quad \dots \quad (2)$$

$$\beta_T = \mu + \mu' \left(\frac{\delta H}{\delta B} \right)_T. \quad \dots \quad (3)$$

With the data obtained from the daily radiosonde flights conducted at the station, values of $(\delta H/\delta B)_p$ and $(\delta H/\delta B)_T$ were found using the least squares method of regression analysis. Over the same period for which β_p and β_T had been calculated, they are as follows:

$$\left(\frac{\delta H}{\delta B}\right)_T = -3.048 \times 10^{-5} \text{ km/mb } (-0.194 \times 10^{-3} \text{ per cent./mb}),$$

$$\left(\frac{\delta H}{\delta B}\right)_p = +1.612 \times 10^{-2} \text{ km/mb } (+0.1028 \text{ per cent./mb}).$$

The mass absorption coefficient μ , the decay coefficient μ' , and the mean range for mesons before decay, L (the reciprocal of μ' , where μ' is expressed as the fractional change of intensity per km, $(1/I)(\delta I/\delta H)$), obtained by substitution in equations (2) and (3), compare favourably with the values found by Trumpy in Norway, as shown in Table 3.

TABLE 3

Place		μ (%/mb)	μ' (%/km)	L (km)
Macquarie Island	..	-0.120	-5.52	18.1
Norway	-0.151	-5.8	17.3

V. CONCLUSIONS

Summing up, it appears that marked changes in cosmic ray intensities occur following the passage of cold fronts, and they become more apparent after correction with the total bi-monthly barometer coefficients. The apparent increases are due to changes in the value of $\delta H/\delta B$, upon which the barometer coefficient depends, $\delta H/\delta B$ having a greater value (in the positive sense) for cold dry air masses than for warm moist air. The frequently observed fluctuations of short-term barometer coefficients may be largely traceable to this cause.

VI. ACKNOWLEDGMENTS

Acknowledgment is due to Drs. A. G. and K. B. Fenton, who suggested and encouraged this work; to Professor A. L. McAulay for his valuable criticisms of the draft paper; to Messrs. Martin and Gibbs of the Melbourne Weather Bureau for extracting the 1950 frontal data; to the A.N.A.R.E. for providing the cosmic ray data; and to the C.S.I.R.O. for the funds which enabled this study to be continued at Hobart.

VII. REFERENCES

- CARO, D. E., LAW, P. G., and RATHGEBER, H. D. (1948).—*Aust. J. Sci. Res.* A 1: 261-74.
 DUPERIER, A. (1949).—*Proc. Phys. Soc. Lond.* A 62: 684-96.
 GIBBS, W. J., GOTLEY, A. V., and MARTIN, A. R. (1952).—*Aust. Nat. Antarctic Res. Expedition Reports*, Series D, Meteorology. Vol. 1, Part 1 (c).
 LINDHOLM, F. (1950).—*Tellus* 2: 63-7.
 LOUGHRIDGE, D. H., and GAST, P. F. (1940).—*Phys. Rev.* 58: 583-5.
 NISHINA, Y., SEKIDO, Y., SIMAMURA, H., and ARAKAWA, H. (1940a).—*Phys. Rev.* 57: 663.
 NISHINA, Y., SEKIDO, Y., SIMAMURA, H., and ARAKAWA, H. (1940b).—*Phys. Rev.* 57: 1050.
 TRUMPY, B. (1949).—*Univ. Bergen Arb.* 1949, No. 3.

AN INVESTIGATION OF FINITE STRAIN IN AN ISOTROPIC MATERIAL SUBJECTED TO HYDROSTATIC PRESSURE AND ITS SEISMOLOGICAL APPLICATIONS

By A. KEANE*

[*Manuscript received December 14, 1953*]

Summary

An attempt is made to reduce the generality of the usual theories of finite strain and to concentrate on the case of hydrostatic pressure. The method used has been based on the works of Murnaghan (1951) and Birch (1952), and yields a general relation giving the elastic constants in terms of the strain and the derivatives of pressure with respect to strain.

It becomes evident that a law in lieu of Hooke's law is needed to proceed with a mathematical theory; and so various elasticity equations in the literature are investigated to determine the special assumptions used to derive them from the present general equations. A linear relation between the second and third order elastic constants is then proposed and the resulting relation between pressure and density compared with Bridgman's experimental results for the alkali metals. It is found that the proposed law of finite hydrostatic strain agrees favourably with experiment and also with deductions from the atomic theory of solids.

The paper concludes with a few relevant remarks on the implications of the present theory to certain seismological problems.

I. INTRODUCTION

The aim of this paper is to investigate the elasticity equations governing large hydrostatic strains, so that the effect of the pressure existing in the deep interior of the Earth can be assessed. Considerable simplification results on treating this special problem instead of developing a general theory.

Following the terminology suggested by Murnaghan (1937), we say that if the displacements are referred to the initial state as standard we are using the Lagrangian approach, while if the final state is taken as standard we have the Eulerian approach. Throughout this paper the Lagrangian viewpoint will be adopted.

In the first part of the paper the assumption is made that the pressure can be expressed as a power series in the strain. This is seen to be equivalent to the usual assumption that the elastic strain energy can be so expanded.

All the equations herein can easily be derived on the Eulerian approach, which apparently has neither advantage nor disadvantage except that, in approximations like that made by Birch (1952), the two methods give widely different results.

* Wollongong Technical College, Wollongong, N.S.W.

II. THE COMPRESSIBILITY RATIO

Since we are concerned with hydrostatic strain we may refer a particle to the same set of Cartesian axes before and after the application of strain. In the following we take the coordinates of a typical particle in the initial state as (a, b, c) and in the final state as (x, y, z) where $(x, y, z) = (a - ka, b - kb, c - kc)$. We denote the displacement of the particle by (u, v, w) . Regarding the medium as isotropic the initial state may be previously subjected to hydrostatic strain, according to the work of Murnaghan (1951, p. 64).

Using the Lagrangian approach the components of the strain tensor for hydrostatic strain are

$$\varepsilon_{aa} = \varepsilon_{bb} = \varepsilon_{cc} = \varepsilon_L, \quad \varepsilon_{bc} = \varepsilon_{ca} = \varepsilon_{ab} = 0,$$

where

$$\varepsilon_{aa} = \frac{\partial u}{\partial a} + \frac{1}{2} \left(\frac{\partial u}{\partial a} \right)^2,$$

so that

$$\varepsilon_L = -k + \frac{1}{2}k^2 \quad \text{or} \quad 1 - k = (1 + 2\varepsilon_L)^{\frac{1}{2}}.$$

If we take V_0 as the volume contained in a rectangular parallelepiped of sides a, b, c , in the initial state and V as the volume in a rectangular parallelepiped of sides x, y, z in the final state, then

$$V = xyz = abc(1 - k)^3 = V_0(1 - k)^3,$$

or

$$\frac{V}{V_0} = (1 + 2\varepsilon_L)^{3/2},$$

where the strain ε_L is intrinsically negative. It is usual to write $\varepsilon_L = -e_L$ so that the equation reduces to

$$\frac{\rho}{\rho_0} = (1 - 2e_L)^{-3/2}, \quad \dots \dots \dots (1)$$

where ρ, ρ_0 are the densities in the final and initial states respectively.

III. THE GENERAL STRESS-STRAIN RELATION

We assume that, for an isotropic medium subjected to an initial hydrostatic pressure p_0 , the total pressure p is given by

$$p = p_0 + A_1\varepsilon + A_2\varepsilon^2 + \dots + A_n\varepsilon^n + \dots, \quad \dots \dots \dots (2)^*$$

where the parameters $A_n = A_n(p_0)$ are functions of the initial pressure and the strain ε is measured from the configuration in which $p = p_0$. Hence if the initial pressure is zero

$$p = A_{01}e + A_{02}e^2 + \dots + A_{0n}e^n + \dots, \quad \dots \dots \dots (3)$$

where the coefficients A_{0n} are constants and the strain e is measured from the configuration in which $p = 0$. Note that in this section the strains ε and e are taken as intrinsically positive, corresponding to equation (1), and the symbols are not to be confused with those of Section II.

* It has been noted by the referee that (2) with coefficients given by (5) follows from (3) and $e = e_0 + (1 - 2e_0)\varepsilon$ by Taylor's theorem.

Let the density when $p=0$, p_0 , p be $\rho=\rho_0$, ρ_{p_0} , ρ respectively. If the strain at pressure $p=p_0$ be measured from the configuration in which the pressure is zero, we have

$$\frac{\rho_{p_0}}{\rho_0} = (1-2e_0)^{-3/2}, \quad \frac{\rho}{\rho_0} = (1-2e)^{-3/2}, \quad \frac{\rho}{\rho_{p_0}} = (1-2\varepsilon)^{-3/2},$$

so that

$$1-2e = (1-2e_0)(1-2\varepsilon)$$

or

$$de = (1-2e_0)d\varepsilon. \quad \dots\dots\dots (4)$$

On differentiating equation (2) n times we obtain

$$\frac{d^n p}{d\varepsilon^n} = n!A_n + (n+1)!A_{n+1}\varepsilon + \dots,$$

which, on using (4), reduces to

$$(1-2e_0)^n \frac{d^n p}{de^n} = n!A_n + (n+1)!A_{n+1}\varepsilon + \dots$$

On taking $\varepsilon=0$ so that $e=e_0$ we have

$$A_n = \frac{(1-2e_0)^n}{n!} \left(\frac{d^n p}{de^n} \right)_{e=e_0},$$

or, regarding A_n as a function of p instead of p_0 (since the initial pressure may be considered as variable), we obtain the relation

$$A_n = \frac{(1-2e)^n}{n!} \frac{d^n p}{de^n}. \quad \dots\dots\dots (5)$$

On differentiating (5) with respect to p , it is easily shown that

$$\frac{dA_n}{dp} = \frac{(n+1)A_{n+1} - 2nA_n}{A_1}, \quad \dots\dots\dots (6)^*$$

* Murnaghan (1951, p. 69) has derived the equation

$$p = p_0 + (3\lambda + 2\mu + p_0)e + \left(3\lambda + 2\mu - 9l - n + \frac{3}{2}p_0 \right) e^2 + \dots$$

from which, on using equation (6) we derive that

$$\frac{d}{dp} \left(\lambda + \frac{2}{3}\mu \right) = \frac{-\frac{1}{3} \left(\lambda + \frac{2}{3}\mu \right) - 2 \left(l + \frac{1}{9}n \right)}{\lambda + \frac{2}{3}\mu + \frac{1}{3}p},$$

which is in conflict with the result obtained by Murnaghan. The slip in Murnaghan's argument is in the change of variable given by the equation

$$\xi = \ln \frac{1}{V},$$

which implies the infinitesimal relation

$$\frac{\delta V}{V} = -\delta\xi + \frac{(\delta\xi)^2}{2!} - \frac{(\delta\xi)^3}{3!} + \dots$$

Defining the incompressibility κ by the relation

$$\kappa = \rho \frac{dp}{d\rho}, \quad \dots\dots\dots (7)$$

it follows from equations (1) and (5) that $A_1 = 3\kappa$.

It becomes evident at this stage that for further advance in the mathematical theory we require $A_1(p)$ as a function of p or $A_2(p)$ as a function of $A_1(p)$ since we can then determine all the other parameters by the use of equation (6). The law of finite hydrostatic strain proposed in the present paper is that $A_2(p)$ should be taken as a linear function of $A_1(p)$. However, we will first investigate the various pressure-density equations found in the literature, in order to determine the relations imposed on $A_1(p)$ and $A_2(p)$ by them.

IV. INCOMPRESSIBILITY A CONSTANT

On taking A_1 constant we find from equation (6) that $A_n = \left(\frac{2^n - 1}{n}\right) A_1$, so that, on substituting in equation (2) and summing the series, we obtain

$$p - p_0 = \kappa \ln \frac{\rho}{\rho_0}. \quad \dots\dots\dots (8)$$

The same relation can be obtained from equation (7), the definition of incompressibility. This equation is equivalent to the finite form of the infinitesimal strain relation and is known to be valid only in the case of small pressure increments.

V. INCOMPRESSIBILITY A LINEAR FUNCTION OF PRESSURE

If we assume $\kappa = \kappa_0 + sp$ we have on integrating the equation defining incompressibility

$$p = \frac{\kappa_0}{s} \left\{ \left(\frac{\rho}{\rho_0} \right)^s - 1 \right\}, \quad \dots\dots\dots (9)$$

where we have taken the initial state as being subjected to zero pressure. We can also obtain this relation by using equation (6) to find A_{0n} in terms of A_{01} and substituting in (3). From (9) we have immediately

$$\kappa = \kappa_0 \left(\frac{\rho}{\rho_0} \right)^s. \quad \dots\dots\dots (10)$$

These equations are those obtained from Murnaghan's (1951) integrated linear theory of elasticity, a method which gives fairly good results for small ranges of pressure, or, as will be shown later, for extremely high pressure ranges. The special case $s = 2/3$ gives the same result as Seth's (1950) method.

VI. A GENERALIZATION OF BIRCH'S EQUATION

From equation (2) we easily derive that

$$p = \frac{1}{(1 - 2\varepsilon)^\beta} \left\{ p_0 + (A_1 - 2\beta p_0)\varepsilon + \left(A_2 - 2\beta A_1 + \frac{\beta(\beta - 1)}{1.2} \cdot 2^2 p_0 \right) \varepsilon^2 + \dots \right\}.$$

Assuming now that the coefficient of ε^2 vanishes, so that

$$A_2 = 2\beta A_1 - \frac{\beta(\beta-1)}{1.2} \cdot 2^2 p_0, \dots\dots\dots (11)$$

it can be demonstrated by the use of equation (6) that all the coefficients of higher powers of ε vanish. Hence

$$p = \frac{1}{(1-2\varepsilon)^\beta} \{p_0 + (A_1 - 2\beta p_0)\varepsilon\},$$

or, if the initial pressure is zero,

$$p = \frac{A_{01}e}{(1-2e)^\beta},$$

where β is given by $A_{02}/2A_{01}$.

In this case the pressure-density relation is

$$p = \frac{3}{2} \kappa_0 \left(\frac{\rho}{\rho_0}\right)^{\frac{2}{3}(\beta-1)} \left\{ \left(\frac{\rho}{\rho_0}\right)^{2/3} - 1 \right\}, \dots\dots\dots (12)$$

and it can be shown that

$$\frac{d\kappa}{dp} = \frac{2\beta^2 \left(\frac{\rho}{\rho_0}\right)^{2/3} - 2(\beta-1)^2}{3\beta \left(\frac{\rho}{\rho_0}\right)^{2/3} - 3(\beta-1)},$$

so that

$$\left. \begin{aligned} \left(\frac{d\kappa}{dp}\right)_{p=0} &= \frac{4\beta-2}{3}, \\ \left(\frac{d\kappa}{dp}\right)_{p \rightarrow \infty} &= \frac{2\beta}{3}. \end{aligned} \right\} \dots\dots\dots (13)$$

The equation (12) reduces to that of Seth when $\beta=1$. Birch (1952), working from a different approach, obtained equation (12) with $\beta=7/2$. This value of β gives reasonable agreement with the results of Bridgman's experiments on the alkali metals but Birch's formula needs modification as will become apparent in Sections VIII and IX. It should be remarked that the equation

$$p = \frac{3}{2} \kappa_0 \left\{ \left(\frac{\rho}{\rho_0}\right)^{7/3} - \left(\frac{\rho}{\rho_0}\right)^{5/3} \right\}, \dots\dots\dots (14)$$

which has been referred to as Birch's equation, is not considered by Birch as an exact relation but merely as the best single-constant formula describing the behaviour of all substances subjected to large hydrostatic strain.

In support of Birch's contention it should be noted that, even if the special relation (11) is not valid, equation (12) is still correct to the second order in e if β is given by $A_{02}/2A_{01}$.

VII. A_2 A LINEAR FUNCTION OF A_1

The present method of approach to the problem of finite hydrostatic strain immediately suggests the assumption $A_2 = \lambda A_1 + \mu$. From equation (6) we obtain

$$\frac{d\kappa}{dp} = \frac{(\alpha - N)\kappa_0 + N\kappa}{\kappa}, \dots\dots\dots (15)$$

where the constants α, N have been introduced as being more convenient than λ, μ . It follows from (15) that

$$\left. \begin{aligned} \left(\frac{dx}{dp} \right)_{p=0} &= \alpha, \\ \left(\frac{dx}{dp} \right)_{p \rightarrow \infty} &= N. \end{aligned} \right\} \dots\dots\dots (16)$$

Integrating equation (15) we obtain

$$p = \frac{\kappa - \kappa_0}{N} - \frac{(\alpha - N)\kappa_0}{N^2} \ln \left\{ 1 + \frac{N(\kappa - \kappa_0)}{\alpha\kappa_0} \right\}. \dots\dots (17)$$

The definition of incompressibility combined with (15) gives

$$\frac{\rho}{\rho_0} = \left\{ 1 + \frac{N(\kappa - \kappa_0)}{\alpha\kappa_0} \right\}^{\frac{1}{N}}, \dots\dots\dots (18)$$

so that (17) reduces to

$$p = \frac{\alpha\kappa_0}{N^2} \left\{ \left(\frac{\rho}{\rho_0} \right)^N - 1 \right\} - \frac{(\alpha - N)\kappa_0}{N} \ln \frac{\rho}{\rho_0}. \dots\dots\dots (19)$$

This equation is a generalization of Murnaghan's equation (9) and reduces to it when $\alpha = N$. The assumption of a linear law is very appealing since it is in the spirit of Hooke's law, and further, the success of Murnaghan's integrated linear theory suggests possibilities in its generalization.*

VIII. COMPARISON OF THEORY WITH EXPERIMENT

Birch (1952) has compared equation (14) with Bridgman's experimentally determined values of p v. v/v_0 for the alkali metals by determining the value of κ_0 for each value of pressure. The results of the calculation are shown in Table 1.

The most noticeable feature of these results is that the calculated value of κ_0 increases with the pressure and over the range under consideration varies by more than 20 per cent. This suggests that Birch's formula will not be useful for extrapolation since κ_0 in this formula varies with pressure. However, as the pressure increases the calculated value of κ_0 becomes nearly constant so that the suggestion is that the formula may be valid at high pressures but is invalid at low pressures. This invalidates its use for extrapolating from densities obtained over a range of high pressures to the density at zero pressure. It further suggests that the value $(dx/dp)_{p=0} = 4$ obtained from the formula needs modification.

The fitting of the experimental data to equation (19) is laborious and difficult since the curve of best fit can only be obtained by a method of trial and error. In the accompanying Tables 2, 3, and 4 are shown examples of

* It is to be noted that the present solution is also a generalization of the results obtained by the writer (1953) as an approximate solution of the equation derived by Murnaghan (1951, p. 70) for dx/dp . In view of the footnote to Section III, it follows that the previous results are in error; but it is evident that the modification to Murnaghan's equation in no way alters the method of approximation used nor does it greatly alter the deductions made therefrom.

curves of best fit to Bridgman's data for lithium, sodium, and potassium. It is significant that the variation of the calculated values of pressure using equation (19) have only half the variation found in the calculated values of Birch. Since Bridgman's two values for the density at 40×10^9 dyn/cm² for sodium corresponds to an error in the calculated pressure of 6 per cent., the greatest error of 7 per cent. in the calculated pressures is almost within the accuracy of the experimental data.

TABLE 1

THE RATIO $2/3\kappa_0$ FOR THE ALKALI METALS CALCULATED FROM EQUATION (14) AND USING BRIDGMAN'S EXPERIMENTAL VALUES (BIRCH 1952)

p (10^9 kg/cm ²)	Lithium		Sodium		Potassium	
	v/v_0	$2/3\kappa_0$	v/v_0	$2/3\kappa_0$	v/v_0	$2/3\kappa_0$
5	0.957	6.41	0.929	11.3	0.884	21.0
10	0.926	5.99	0.883	10.7	0.817	20.2
15	0.899	5.84	0.852	9.8	0.770	19.6
20	0.875	5.81	0.818	10.0	0.732	19.4
25	0.855	5.73	0.791	10.0	0.699	19.6
30	0.835	5.77	0.767	10.0	0.671	19.7
35	0.816	5.81	0.746	10.0	0.647	19.9
40	0.798	5.91	0.727	10.1	0.627	19.9
45	0.782	5.97	0.710	10.1	0.604	20.6
30	0.833	5.84	0.770	9.81	0.668	20.2
40	0.801	5.76	0.737	9.38	0.628	19.7
50	0.773	5.76	0.708	9.21	0.595	19.7
60	0.748	5.78	0.683	9.11	0.568	19.6
70	0.727	5.76	0.661	9.05	0.546	19.5
80	0.707	5.79	0.641	9.05	0.528	19.2
90	0.689	5.83	0.623	9.07	0.513	19.0
100	0.672	5.88	0.606	9.13	0.500	18.7

The values obtained for the density at 45×10^9 dyn/cm² in 1938 for lithium, sodium, and potassium are seen to be unreasonably high, and it is perhaps safe to say that the high values in the 1948 table are not very accurate. This is pertinent since the error in the present calculated values is 4 per cent. for 100×10^9 dyn/cm².

Apart from the variation at 100×10^9 dyn/cm², the greatest error in the tables is for sodium at 15×10^9 dyn/cm². It can be seen from differencing the values of v/v_0 that the experimental value at 15×10^9 dyn/cm² is too low; this conclusion also follows from Birch's calculations of κ_0 .

In view of the above discussion it would appear that formula (19) fits the experimental data to within 4 per cent., which corresponds to a range of less than 8 per cent., while Birch's formula gives a range of from 10 to 20 per cent. in the calculated values of κ_0 .

Murnaghan (1951, p. 78) has compared his second-order theory with experiment and shown that the calculated values of v/v_0 agree with Bridgman's experimentally determined values within 1 per cent. It can be seen from Table 3 for the two values of v/v_0 at 40×10^9 dyn/cm² that an error of 1 per cent. in v/v_0 is equivalent to an error of 5.5 per cent. in p . Similarly from Table 1 it can

TABLE 2
COMPARISON OF EQUATION (19) WITH BRIDGMAN'S EXPERIMENTAL
RESULTS FOR LITHIUM

v/v_0	p (Bridgman) (10^9 dyn/cm ²)	p (calc.)* (10^9 dyn/cm ²)
<i>Bridgman's 1938 values</i>		
0.957	5	5.2
0.926	10	9.9
0.899	15	14.6
0.875	20	19.6
0.855	25	24.2
0.835	30	29.4
0.816	35	34.9
0.798	40	40.7
0.782	45	46.5
<i>Bridgman's 1948 values</i>		
0.833	30	30.0
0.801	40	39.7
0.773	50	49.9
0.748	60	60.4
0.727	70	70.6
0.707	80	81.5
0.689	90	92.5
0.672	100	104.0

* Pressure calculated from the formula

$$p = 79.8 \left\{ \left(\frac{\rho}{\rho_0} \right)^{2.6} - 1 \right\} - 101.6 \ln \frac{\rho}{\rho_0}$$

so that $\kappa_0 = 10.6 \times 10^{10}$ dyn/cm², and $\alpha = 5.1$.

be seen that an error of 1 per cent. in v/v_0 corresponds to an error of 5.5 per cent. in κ_0 . It thus appears that Murnaghan's variation of 2 per cent. in the calculated values of v/v_0 corresponds to a variation of 11 per cent. in either the pressure or the initial incompressibility.

IX. DEDUCTIONS FROM ATOMIC THEORY

The atomic theory of solids (Seitz 1940 ; Fürth 1944) gives some important evidence on the compression of materials at high pressures. In its simplest form the theory will give only indicative results and cannot be considered as rigorous.

Assuming inverse laws of attraction and repulsion we have

$$w(x) = -\frac{a}{x^m} + \frac{b}{x^n}$$

TABLE 3
COMPARISON OF EQUATION (19) WITH BRIDGMAN'S EXPERIMENTAL
RESULTS FOR SODIUM

v/v_0	p (Bridgman) (10^9 dyn/cm ²)	p (calc.)* (10^9 dyn/cm ²)
<i>Bridgman's 1938 values</i>		
0.929	5	5.0
0.883	10	9.8
0.852	15	14.0
0.818	20	19.6
0.791	25	25.0
0.767	30	30.6
0.746	35	36.4
0.727	40	42.3
0.710	45	48.1
<i>Bridgman's 1948 values</i>		
0.770	30	29.9
0.737	40	39.1
0.708	50	48.9
0.683	60	59.1
0.661	70	69.5
0.641	80	80.5
0.623	90	91.8
0.606	100	103.9

* Pressure calculated from the formula

$$p = 38.5 \left\{ \left(\frac{\rho}{\rho_0} \right)^3 - 1 \right\} - 61 \ln \frac{\rho}{\rho_0}$$

so that $\kappa_0 = 5.45 \times 10^{10}$ dyn/cm² and $\alpha = 6.3$.

as the potential between two atoms from which it can be derived that at the absolute zero of temperature

$$\frac{d\kappa}{dp} = \frac{\theta^2 \left(\frac{\rho}{\rho_0} \right)^\theta - \varphi^2 \left(\frac{\rho}{\rho_0} \right)^\varphi}{\theta \left(\frac{\rho}{\rho_0} \right)^\theta - \varphi \left(\frac{\rho}{\rho_0} \right)^\varphi},$$

where $\theta = n/3 + 1$, $\varphi = m/3 + 1$.

Thus as

$$\rho \rightarrow \rho_0, \quad \frac{d\kappa}{dp} \rightarrow \theta + \varphi,$$

$$\rho \rightarrow \infty, \quad \frac{d\kappa}{dp} \rightarrow \theta.$$

The significant deductions to be made from these results are that

$$\left(\frac{d\kappa}{dp}\right)_{p=0} < 2 \left(\frac{d\kappa}{dp}\right)_{p \rightarrow \infty},$$

and

$$\left(\frac{d\kappa}{dp}\right)_{p=0} > \left(\frac{d\kappa}{dp}\right)_{p \rightarrow \infty} + 1.$$

TABLE 4
COMPARISON OF EQUATION (19) WITH BRIDGMAN'S EXPERIMENTAL
RESULTS FOR POTASSIUM

v/v_0	p (Bridgman) (10^9 dyn/cm ²)	p (calc.)* (10^9 dyn/cm ²)
<i>Bridgman's 1938 values</i>		
0.884	5	5.0
0.817	10	10.0
0.770	15	14.6
0.732	20	19.5
0.699	25	24.7
0.671	30	30.2
0.647	35	35.6
0.627	40	41.0
0.604	45	48.0
<i>Bridgman's 1948 values</i>		
0.668	30	30.8
0.628	40	40.7
0.595	50	51.1
0.568	60	61.7
0.546	70	72.0
0.528	80	81.9
0.513	90	91.3
0.500	100	100.2

* Pressure calculated from the formula

$$p = 16 \left\{ \left(\frac{\rho}{\rho_0} \right)^3 - 1 \right\} - 17 \ln \frac{\rho}{\rho_0}$$

so that $\kappa_0 = 3.1 \times 10^{10}$ dyn/cm² and $\alpha = 4.65$.

The first result follows from the fact that $n > m$ since otherwise the crystal would collapse, and the second follows since $\varphi = m/3 + 1$.

It has been shown (Seitz 1940) that for lithium fluoride $n = 6$ so that $\theta = 3$, and further it has been determined that θ increases with increasing atomic weight. It appears that the values obtained for the alkali metals in Tables 2, 3, and 4 are not inconsistent with these results.

Equation (14) is seen to correspond to a law of force between the atoms in the alkali crystal lattice given by

$$w(x) = -\frac{a}{x^2} + \frac{b}{x^4}.$$

Apart from the fact that such a law is only a first approximation, the value $n=4$ is somewhat too low and its assumed constancy for the alkali metals is inconsistent with other evidence. It follows that Birch's formula for high pressures should best fit lithium and become increasingly less accurate for sodium and potassium. This is supported by the results of Table 1. Birch (1952, p. 252) has considered other aspects of the problem.

The constancy of κ_0 as calculated from Birch's formula for rubidium at high pressures is in conflict with the present arguments. However, due to an internal variation of more than 12 per cent. in the experimental data, it appears that no definite deductions can be made. Another indication of incompatibility is the large value of κ_0 , which is nearly equal to the value found for potassium.

Within the accuracy of the experimental data the results for the alkali metals obtained on using equation (17) satisfy the following three conditions imposed by the atomic theory of solids.

- (i) $\left(\frac{d\kappa}{dp}\right)_{p=0} < 2\left(\frac{d\kappa}{dp}\right)_{p \rightarrow \infty},$
- (ii) $\left(\frac{d\kappa}{dp}\right)_{p=0} > \left(\frac{d\kappa}{dp}\right)_{p \rightarrow \infty} + 1,$
- (iii) $\left(\frac{d\kappa}{dp}\right)_{p \rightarrow \infty}$ increases with atomic weight.

X. SEISMOLOGICAL DEDUCTIONS

(1) For pressures above $0.5\kappa_0$ the value of $d\kappa/dp$ is somewhat constant. This is in agreement with deductions made by Ramsey (1950) from atomic theory. Graphs of incompressibility against depth and of pressure against depth obtained by Bullen (1947, p. 220) for the interior of the Earth lead to the inference that κ is not a linear function of p in the Earth's inner core and further that $d\kappa/dp$ increases with pressure. Both these facts suggest that Bullen's values need modification in this region. Perhaps the assumed constancy of $d\kappa/dp$ would give the calculations for the inner core a greater degree of determinateness.

(2) On extrapolating the ratio $\varphi = \kappa/\rho$ for layer D of the Earth to zero pressure, Birch (1952) has found that the value of φ_0 is too high to agree with the experimentally determined values for any of the most likely components of this layer. On the basis of the discussion given earlier it would appear that on extrapolating with Birch's formula the value of κ_0 would be too high while at the same time ρ_0 would be too low. Thus on using Birch's formula the extrapolated value of φ_0 would be increased markedly from its true value.

(3) Figure 1 shows the variation of κ/ρ with pressure for potassium. The large initial increase may help to explain part of the large increase in seismic velocities in layer C of the Earth.

(4) Both equations (12) and (19) when fitted to Bridgman's results give values of $(d\kappa/dp)_{p \rightarrow \infty}$ greater than $5/3$. This is in conflict with Elsasser's (1951)

smooth interpolation of Bridgman's experimental results to those of quantum mechanics and Bullen's (1952) caution, that considerable care should be shown in any such interpolation, must be supported.

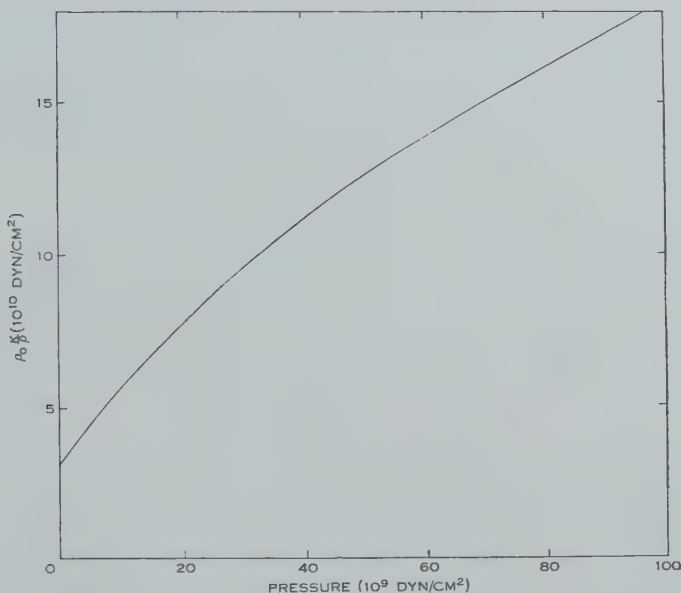


Fig. 1.—Variation of κ/ρ with pressure for potassium.

XI. ACKNOWLEDGMENTS

Professor G. Bosson of the New South Wales University of Technology has read through the manuscript at various stages of its development and made many helpful suggestions. For this assistance the author wishes to express his appreciation. He is also indebted to Professor F. D. Murnaghan, previously of the Instituto Tecnológico de Aeronautica, Brazil, and Professor Francis Birch of Harvard University, U.S.A., for helpful replies to some of his queries.

XII. REFERENCES

- BIRCH, F. (1952).—Elasticity and constitution of the Earth's interior. *J. Geophys. Res.* **57**: 227-86.
- BULLEN, K. E. (1947).—“An Introduction to the Theory of Seismology.” (Cambridge Univ. Press.)
- BULLEN, K. E. (1952).—On density and compressibility up to thirty million atmospheres. *Mon. Not. R. Astr. Soc. Geophys. Suppl.* **6**: 383-401.
- ELSASSER, W. M. (1951).—Quantum-theoretical densities of solids at extreme compression. *Science* **113**: 105-7.
- FÜRTH, R. (1944).—On the equation of state for solids. *Proc. Roy. Soc. A* **183**: 87-110.
- KEANE, A. (1953).—Variation of the incompressibility of an elastic material subjected to large hydrostatic pressure. *Nature* **172**: 117.
- MURNAGHAN, F. D. (1937).—Finite deformations of an elastic solid. *Amer. J. Math.* **59**: 235-60.
- MURNAGHAN, F. D. (1951).—“Finite Deformation of an Elastic Solid.” (Wiley: New York.)
- RAMSEY, W. H. (1950).—On the compressibility of the Earth. *Mon. Not. R. Astr. Soc. Geophys. Suppl.* **6**: 42-9.
- SEITZ, F. (1940).—“The Modern Theory of Solids.” (McGraw-Hill: New York.)
- SETH, B. R. (1950).—Finite elastic deformations. *Proc. Symp. Appl. Math.* **3**: 67-84.

ON THE SECULAR VARIATION OF RAINFALL AT ADELAIDE

By E. A. CORNISH*

[*Manuscript received January 25, 1954*]

Summary

A detailed analysis of the rainfall of Adelaide has established that periodic changes occur in the incidence and duration of the winter rains. These changes have a period and amplitude of approximately 23 years and 30 days respectively, and superimposed on them is a long-term trend which is manifested by protraction of the latter half of the season, spring rains now occurring about 3 weeks later than they did just over 100 years ago. The total quantity of rain precipitated has shown no statistically significant changes.

I. INTRODUCTION

The daily rainfall observations recorded at Adelaide during the 95 years 1839–1933 inclusive have been subjected to detailed analysis, and the results reported (Cornish 1936). This investigation pointed to the occurrence of periodic changes in the incidence and duration of the winter rains, with a period of approximately 23 years and amplitude of 30 days, but no statistically significant changes were demonstrable in the amount of rain precipitated. The periodic trend was very clearly defined from 1839 to about 1912, but thereafter the observations showed some evidence of a breakdown in the law of change. With the accumulation of further data, the opportunity has been taken to re-examine this point and to revise the analysis.

II. DATA

On January 1, 1839, Sir George Kingston established a daily rainfall record at Adelaide on a site approximately 500 yd from the present position of the Observatory. This record was continued until November 1879. From May 1860, readings have been taken at the Observatory, so that over 19 years the two sets of observations were concurrent. During this interval, the average annual difference between the gauges was only 0.26 in. (Kingston's being the greater), and, considering the close proximity of the sites, it may be assumed that the two series in combination give a continuous and practically uniform record of the Adelaide rainfall. No definite statement could be found regarding the diameter of the gauge employed by Kingston, nor, in fact, the size of the gauge used in the early days of the existence of the Observatory. It is fairly certain, however, that no radical departure could have been made from the standard 8-in. gauge which has been in use since 1870.

The analysis previously reported has now been extended to include the rainfall observations of 1934–1950 inclusive.

* Section of Mathematical Statistics, C.S.I.R.O., Adelaide.

III. ANALYSIS

The basic data were obtained by a method devised by Fisher (1924). The rainfall of each year was divided into 61 six-day totals, and to each of these annual sequences a series of orthogonal polynomial functions of the fifth degree in time was fitted, thus furnishing six quantities with which the amount and distribution of rain in each year could be represented. These distribution constants, designated a' , b' , . . . , f' , have been tabled (Cornish loc. cit.) for the years 1839–1933 inclusive, and supplementary data for the period 1934–1950

TABLE I
RAINFALL DISTRIBUTION VALUES FOR 1934–1950
Unit: 10^{-3} in.

Year	a'	b'	c'	d'	e'	f'
1934	332	+96	—27	—47	—16	+ 3
1935	384	+23	—64	— 6	+17	— 5
1936	317	+44	— 1	+ 1	+31	—11
1937	378	+15	—30	— 7	+18	+ 1
1938	316	—46	—49	+26	+ 5	— 1
1939	381	— 6	—48	— 9	— 1	—12
1940	265	+ 9	—23	— 2	+16	—16
1941	370	—13	—26	—42	— 1	+27
1942	417	+ 6	—94	— 9	+27	— 9
1943	291	—27	—27	— 5	— 1	+13
1944	281	+20	—18	+33	+ 2	—18
1945	293	+64	— 9	—26	— 4	— 2
1946	370	—23	—11	+17	+10	+13
1947	359	+41	—47	— 3	+ 6	+16
1948	351	+33	—51	+15	—14	—23
1949	299	+43	—29	—29	—27	+ 1
1950	263	+29	—52	—10	+11	— 9
Mean 1839–1950	345.35	+14.80	—56.07	— 2.72	+15.67	— 3.28

inclusive are given in Table 1. For convenience of presentation in tabular form, the unit is 10^{-3} in. These quantities are actually proportional to the coefficients of the polynomial terms, and to obtain the latter they must be divided by factors of the form

$$\{(r!)^{260.59} \dots (61-r)\}/(2r+1)!,$$

where r is the degree of the term fitted.

The first constant a' represents the average rainfall in 10^{-3} in. per 6-day period, b' is proportional to the linear term of the seasonal sequence, c' is proportional to the parabolic term, and so on, taking more complex features of the seasonal distribution into account.

In Figures 1, 2, and 3, the courses of the secular changes in a' , b' , and c' respectively have been depicted by plotting running 10-year means of each.

The diagram of a' shows quite marked changes in the amount of rain precipitated, with two short periods of high rainfall centred at 1850 and 1920, when the mean annual rainfall was approximately 24 in., separated by a long interval in which the mean fell to about 20 in.

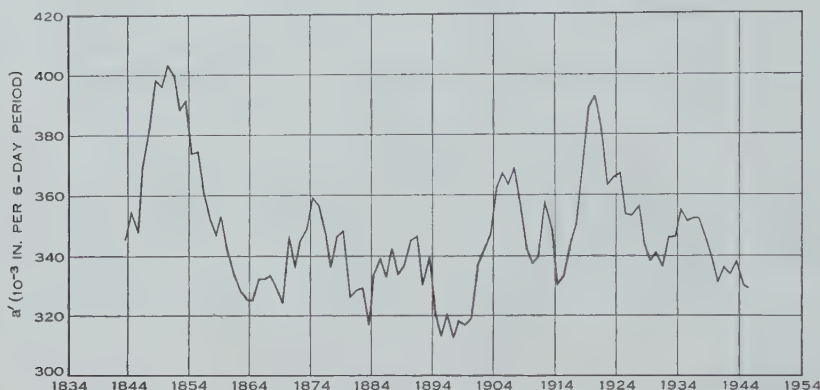


Fig. 1.—Ten-year means of the distribution value a' .

Changes in the course of b' are also prominent, and even more complex than those observed with a' , the diagram showing clearly that the swing in the mean is again becoming regular, following the departure observed between 1910 and 1930. As will appear later, the effects of this disturbance are not so serious as

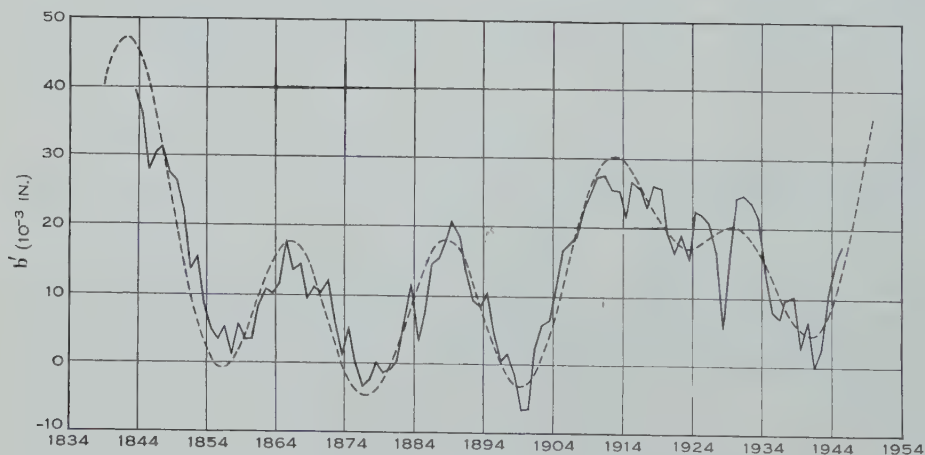


Fig. 2.—Ten-year means and harmonic curve of the distribution value b' .

the figure might indicate. By comparison, c' follows a simple course with a distinct and apparently uniform increase, the mean having increased from -67 in 1839 to -45 in 1950 (both values approximate).

The secular changes observed in the six distribution constants were then examined in more detail by fitting each series of 112 values with an orthogonal

polynomial of the fifth degree in time, and the significance of the polynomials re-tested on all the available data. The procedure adopted in this step may be illustrated by the calculations relating to c' . The following values were obtained, the unit being 10^{-3} in. :

Mean	-56.07143	x'_1	-593.40	$(x'_1)^2$	352,123.56
Coefficient of 1st degree	3.67983	x'_2	68.06	$(x'_2)^2$	4,632.16
" " 2nd "	1.76860	x'_3	42.99	$(x'_3)^2$	1,848.14
" " 3rd "	0.25608	x'_4	7.56	$(x'_4)^2$	57.15
" " 4th "	1.03557	x'_5	35.95	$(x'_5)^2$	1,292.40
" " 5th "	-1.07097	x'_6	-42.98	$(x'_6)^2$	1,847.28

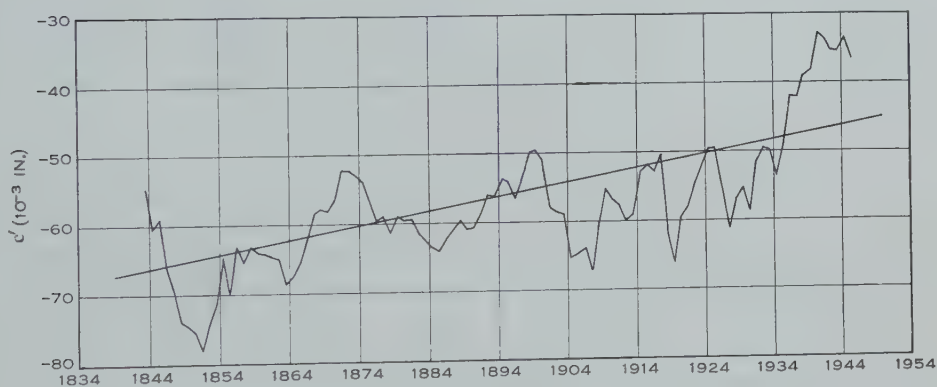


Fig. 3.—Ten-year means and linear regression of the distribution value c' .

The quantities x'_1, \dots, x'_6 are orthogonal and normal linear functions of the 112 values of c' , obtained from the corresponding quantities in the first column by multiplying the coefficient of the term of degree r by

$$\left\{ \frac{(2r+1).112.113 \dots (112+r)}{111.110 \dots (112-r)} \right\}^{\frac{1}{2}},$$

and x'_2, \dots, x'_6 represent the several components of secular change in the c' sequence. The last column gives the squares of these functions. The total variation of the c' values from their mean may be divided into two portions :

- (1) a sum of squares associated with regression on time, and due to a comparatively simple temporal trend predominating over the random fluctuations,
- (2) the remaining sum of squares which may be attributed to random annual variation,

and the first of these components may be partitioned further to give the individual contributions of the terms of the several degrees to the total for regression. The analysis of variance given in Table 2 establishes the strong significance ($P < 0.01$) of the linear term, thus confirming the feature noted in Figure 3.

Table 3 provides a summary analysis of all six distribution constants, the unit being 10^{-3} in.

No significant changes have occurred in a' , d' , e' , and f' , which thus merely fluctuate with large standard deviations about their respective mean values. This confirms the previous findings with respect to a' , d' , and e' . With f' , however, the original analysis indicated an apparently significant parabolic

TABLE 2
ANALYSIS OF VARIANCE OF THE c' SERIES

Variation Due to	Degrees of Freedom	Sum of Squares	Mean Square	Variance Ratio
Linear component	1	4,632.16	4,632.16	6.94
Quadratic „	1	1,848.14	1,848.14	n.s.
Cubic „	1	57.15	57.15	n.s.
Quartic „	1	1,292.40	1,292.40	n.s.
Quintic „	1	1,847.28	1,847.28	n.s.
Deviations from regression ..	106	70,714.30	667.12	
Total	111	80,391.43		

component, but this result had been treated with caution since the f' series had shown a considerable departure from normality. It is of some interest to see that, with more extensive data, this term now degenerates to non-significance.

The significance of the linear term in the course of c' is a new feature, and the remaining significant effects, namely, the quadratic and cubic components

TABLE 3
SUMMARY POLYNOMIAL ANALYSIS OF RAINFALL DISTRIBUTION VALUES
Unit: 10^{-3} in.

	a'	b'	c'	d'	e'	f'
Mean	345.35	14.80	-56.07	-2.72	15.67	-3.28
x_2'	-45.22	-7.49	68.06	-16.98	-30.87	15.94
x_3'	17.46	62.70	42.99	-12.63	-23.20	22.72
x_4'	-72.95	-64.85	7.56	14.91	-32.91	-27.55
x_5'	-116.24	36.96	35.95	3.02	-21.84	-6.49
x_6'	65.14	19.22	-42.98	11.88	1.20	-16.77
Standard residual ..	68.36	29.88	25.83	18.91	17.55	14.10
t (106 deg. freedom).		5.24	22.97	1.52	9.45	2.46

of the changes in b' , confirm the conclusion drawn earlier. The analysis of the b' sequence was actually carried as far as the term of the ninth degree, but since coefficients of terms beyond the fifth degree are quite insignificant they have not been quoted.

Oscillations in the course of b' not associated with the polynomial were conspicuous, and it thus appeared desirable to test whether a harmonic series

would take account of them. The analysis was taken as far as the fifth harmonic, yielding the series

$$14.80 + 6.72 \cos \theta - 5.00 \sin \theta + 2.46 \cos 2\theta + 4.68 \sin 2\theta + 6.96 \cos 3\theta \\ + 5.02 \sin 3\theta + 4.89 \cos 4\theta - 0.12 \sin 4\theta + 3.96 \cos 5\theta + 10.03 \sin 5\theta,$$

where $\theta=0$ in 1839. The significance of this series is due to the fifth harmonic. The fitted curve is given in Figure 2, and evidently follows the oscillations not represented by the polynomial. For this reason, the harmonic series was used in place of the polynomial fitted to b' in subsequent statistical tests.

An attempt was made to provide a more precise test of the significance of the harmonic series fitted to the b' sequence and the polynomials fitted to the remaining distribution constants, by subjecting the whole of the original data, comprising 6832(112×61) six-day totals of rainfall to one comprehensive analysis of variance. With rainfall measured in inches, the primary subdivision of the total variation is as follows:

Variation Due to	Degrees of Freedom	Sum of Squares
Between 6-day means ..	60	159.3083
Between years	111	31.7711
Residual	6660	1412.7703
Total	6831	1603.8497

In the further subdivision of these portions of the total variation, to show the contributions of the harmonic and the polynomials, the individual sums of squares must be placed on a comparable basis. This is accomplished by normalizing the distribution constant corresponding to the original polynomial of degree r using the factor

$$\left\{ \frac{(2r+1).61.62 \dots (61+r)}{60.59 \dots (61-r)} \right\}^{\frac{1}{2}},$$

or, more simply, by multiplying the appropriate sum of squares in the analysis, by the square of the above quantity. This test confirmed the significances obtained above, but, in the presence of such excessive annual variation, is still not sensitive enough to detect any other secular effects.

The last line of Table 3 gives the results of tests of significance on the mean values of the constants, with the exception of a' . Apart from the mean value of d' , the remainder are significant and thus may be regarded as permanent characteristics of the season. The average seasonal distribution, represented by these mean values, is given in Figure 4, where it is contrasted with the means of the 6-day totals (on a per-day basis) and the distribution obtained by fitting the average 6-day totals with a harmonic series which takes the form

$$5.77 - 3.38 \cos \theta - 0.384 \sin \theta + 0.290 \cos 2\theta - 0.636 \sin 2\theta,$$

where $\theta=0$ at the mid point of the first 6-day interval in the year. This curve is identical with what would have been obtained if harmonics of this form had been fitted to the 61 values of each year, and averaged over the 112 years. Further discussion of the polynomial and harmonic representations is given below.

IV. REPRESENTATION OF THE SEASON

When the original investigation was reported (Cornish loc. cit.), Whipple (1936) criticized the representation of individual years of rainfall and the average annual seasonal distribution by means of polynomials. The polynomials were designed to represent individual seasons, and are appropriate for this purpose because they do not impose the condition of periodicity on the weather, which does not, in fact, repeat itself year by year. On the other hand, the *average* rainfall sequence, being periodic, is properly represented by a harmonic curve.

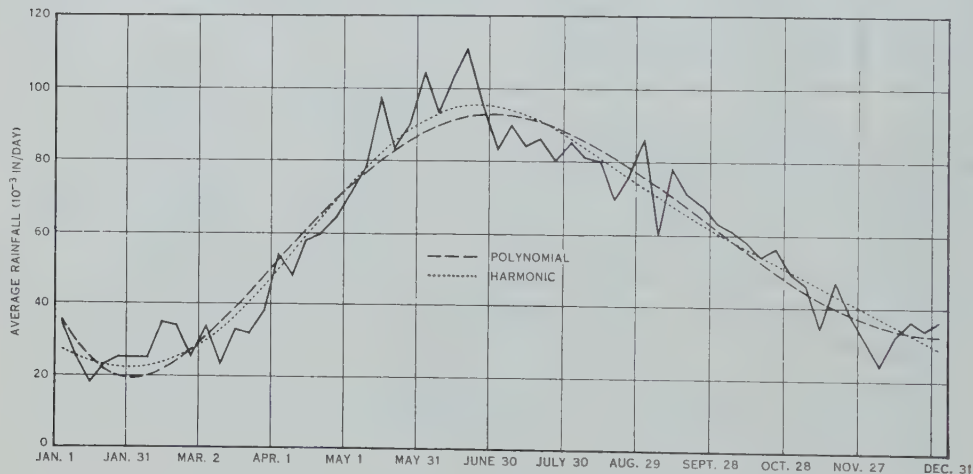


Fig. 4.—Seasonal variation of average daily rainfall.

As polynomials had been used for the annual observations, it was a natural step to show how they represented the average season, but at no stage was it claimed that they were designed for this purpose, or that their average was superior to a harmonic series. Whipple stated "the 12 monthly totals would give twice as much information about the distribution of rain through the year as six terms of the series of polynomials", but gave no reason why they should do so when fitted with a harmonic series.

Instead, however, of using mean monthly rainfall, advantage may be taken of the finer detail supplied by the 61 six-day totals, since the superiority of either method of representation will be demonstrated more clearly with these totals than with the monthly means. The primary subdivision of the sum of squares given in Section III forms the starting point. The contribution made by the average polynomial is 145.8501 , and the sum due to the harmonic is 148.2759 , so that, after making allowance for the secular changes in the distribution

constants, the two curves can be properly contrasted in the following analysis of variance :

Variation	Polynomial Curve			Variation	Harmonic Curve		
	Degrees of Freedom	Sum of Squares	Mean Square		Degrees of Freedom	Sum of Squares	Mean Square
Polynomial ..	5	145·8501	29·1700	Harmonic ..	4	148·2759	37·0690
Residual ..	55	13·4582	0·2447	Residual ..	56	11·0324	0·1970
Between 6-day means ..	60	159·3083			60	159·3083	
Residual ..	6630		0·2114				

Neither mean square, 0·2447 and 0·1970, differs significantly from 0·2114, and consequently both curves represent the average sequence well. The harmonic gives slightly the better representation, since its residual variance is smaller (and it absorbs one less degree of freedom), but comparison of the respective mean squares does not substantiate Whipple's estimate. The amount of information is the inverse of the variance, and the figures are 4·09 units for the polynomial and 5·08 units for the harmonic, the latter being only 25 per cent. greater. It must be emphasized again that the test has been made merely to examine the force of Whipple's statement.

TABLE 4
TEST OF NORMALITY OF RAINFALL DISTRIBUTION VALUES

Measure of Departure	a'	b'	c'	d'	e'	f'	Standard Deviation
g_1	0·2873	0·1666	0·0400	0·0781	0·2398	0·3900	0·2284
g_2	-0·3400	-0·1270	-0·2474	0·3485	0·4626	2·6497	0·4531

V. FREQUENCY DISTRIBUTION OF a' , b' , . . . , f'

All tests of significance employed are based upon the assumption of normality of the distribution constants a' , b' , . . . , f' . Table 4 summarizes a test of normality, and shows the measures of departure with standard deviations appropriate to a normal distribution. These tests confirm the results previously given, f' again being the only quantity to show a significant departure from normality ; the positive value of g_2 indicates a symmetrical departure such that the apex and two tails of the distribution are increased at the expense of the shoulders.

VI. CORRELATIONS OF THE DISTRIBUTION CONSTANTS

It has been observed that in the South Australian environment intra-station correlations of rainfall for subdivisions of the season are very weak, and in this connexion it is of some interest to compare these observations with correlations

obtained from the extensive Adelaide record. The distribution constants have been calculated from uncorrelated functions of time, so that if rainfall at any time of the season is correlated with rainfall at any other time, the relation should appear on correlating them. After making allowance for secular changes, the correlations were determined and their inverse hyperbolic transformations $z = \tanh^{-1} r$, where r is the correlation coefficient, are presented in Table 5. The variate z is very nearly normally distributed, each value in the table having a standard deviation of 0.0958. Judging the data as a whole, the covariation is undoubtedly real ($\chi^2 = 81.35$ with 15 degrees of freedom), but obviously from the individual values of z the correlations are all very weak, a result which confirms the observations made for stations distributed throughout the winter rainfall zone in South Australia.

TABLE 5
INTER-CORRELATIONS OF RAINFALL DISTRIBUTION VALUES
Values of $z = \tanh^{-1} r$: standard deviation 0.0958

	a'	b'	c'	d'	e'
b'	0.02536				
c'	-0.5565	0.02709			
d'	0.006209	-0.2769	0.06659		
e'	0.2420	0.1539	-0.2536	-0.03694	
f'	-0.2665	0.1121	0.2785	-0.06205	-0.2030

VII. CHANGES IN THE RAINFALL SEQUENCE

To express in simpler form the changes in the rainfall sequence represented by the secular trends of b' and c' , the data were next presented from another standpoint. Examination of the monthly records disclosed that the main fluctuation seemed to concern the date of incidence of the winter rains, and it therefore appeared desirable to ascertain to what extent the oscillation in this date and the mean date of attainment of the yearly maximum of rainfall would account for the disturbances observed. In addition, there was definite evidence that spring and early summer rains had advanced toward the end of the year.

Consider the total rainfall of any two successive periods of 183 days. If the date dividing them falls in the spring, the second will generally contain the smaller quantity of rain, and vice versa if the day of division falls in the autumn. When daily differences of such totals are taken, a sequence of values should be obtained which changes sign regularly twice per rainfall year, first in the winter from negative to positive, and secondly in the summer from positive to negative.

These differences were determined for each day of the year in the 112 years of observations. With the exception of several years, the change of sign in winter and summer was defined very clearly. In exceptional years, the differences alternated in sign for short periods of varying lengths before definitely adopting the opposite sign, and in such cases the day of zero difference was obtained by smoothing the series with successive 10-day means.

The date in the winter at which the preceding 6 months had received as much rain as the 6 months following may be regarded as an empirical median of the rainfall sequence, and, in the original investigation of the changes in the mean value of b' , two quartiles were located, one on either side of the median, between each of which and the median, one-quarter of the rainfall of the year surrounding the median date, had fallen. To these dates, three others corresponding to the octiles $\frac{3}{8}$, $\frac{5}{8}$, and $\frac{7}{8}$ have been added. These additional figures

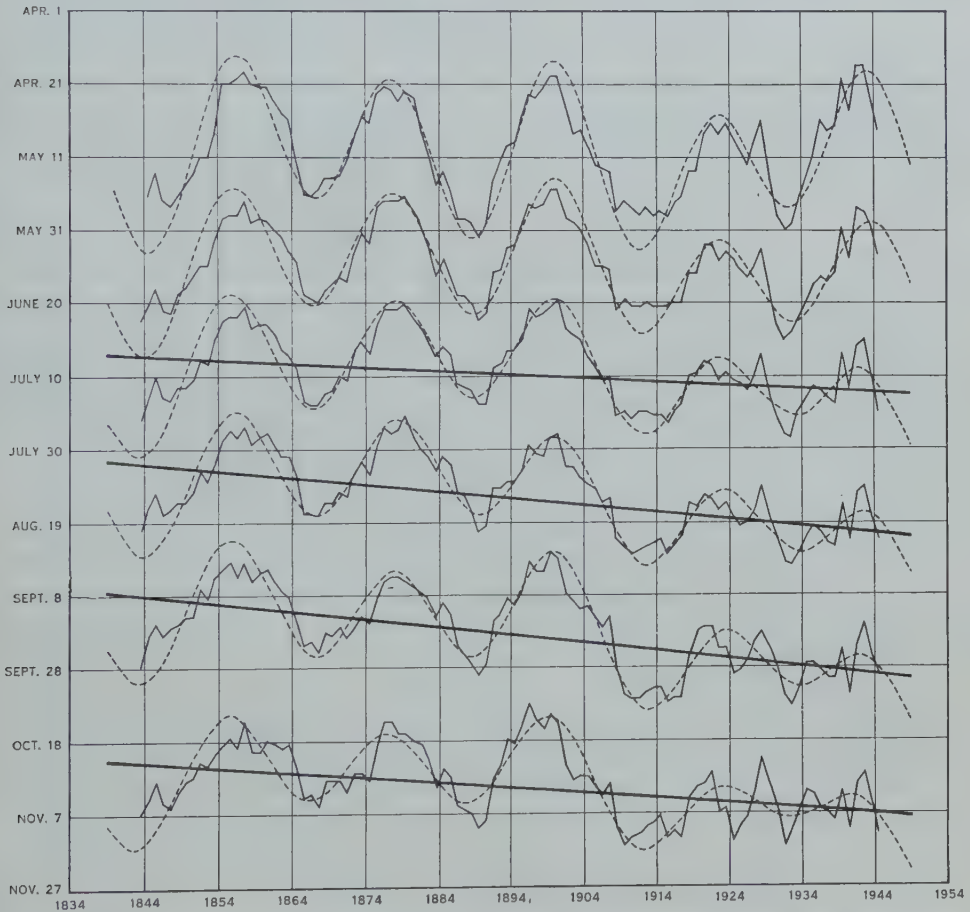


Fig. 5.—Ten-year means and harmonic curves of the octiles of the rainfall year, with linear components in the dates from the median onward.

serve to provide more detail of the changes expressed by b' , but their principal use is to illustrate more clearly the linear trend represented by c' . The whole series is given in Figure 5 by plotting 10-year means of the several dates, and superimposing the harmonic curves *fitted to the original data*.

The six graphs all agree with the course of b' . The periods closely approximate each other, and the movements are in phase, but the displacements are in opposite directions, which must necessarily follow from the physical

nature of the coefficient b' . The abnormal behaviour of the latter between 1910 and 1930 is reflected in all the curves of Figure 5, though not nearly to the same degree in the first quartile and third octile.* The reason for this is that the curves of Figure 5 are based on the rainfall year, which is a natural unit, whereas b' is dependent upon rain falling in the calendar year. There is definite evidence also from the curves of the first quartile and third octile that the course of b' is again becoming regular.

The changes expressed by b' are quite distinct from those of c' , and represent a regular oscillation, of which the period and amplitude are approximately 23 years and 30 days respectively, in the dates of incidence of the winter rains and attainment of the yearly maximum of rainfall. This oscillation has been confined principally to a portion of the seasonal distribution, for the date of the minimum has remained practically constant throughout the 112 years, the section exhibiting greatest movement being that extending from April to November.

The amplitude of cycles occurring after 1910 is gradually reduced in the curves from the median date onward, and in the seventh octile has been almost entirely eliminated. This is undoubtedly due to the progressive change represented by c' ; evidently some considerable time had to lapse before the cumulative effect became dominant, because the first three oscillations are clearly defined. The linear components in the curves from the median onward are all strongly significant, and are given in Figure 5. As the diagrams indicate, the changes expressed by c' have been confined to the latter half of the season, and amount to protraction of the later rains, the median, fifth octile, third quartile, and seventh octile having advanced by 11, 21, 24, and 17 days respectively in the period under review. The record is too short to permit any definite statement regarding this remarkable feature of the data, but it probably constitutes portion of a long-term oscillation, of which all that can be said at the moment is that the period (if it is periodic) is not less than 225 years.

The mutual agreement of the curves substantiates the view that a shift of date, regardless of quantity of seasonal precipitation, provides an adequate description of the changes in progress. In absolute terms, neither the amplitude of the cyclic movement, nor the total extension of the season, is great, but relative to a season of short duration such as occurs over a large proportion of the winter rainfall zone in South Australia, they may well assume considerable significance, particularly from the standpoint of agriculture.

VIII. CONCLUDING REMARKS

Whipple (*loc. cit.*) was the first to draw attention, in publication, to the extraordinary coincidence of this oscillation in the rainy season, with the alternation, from cycle to cycle, of the magnetic polarity of sunspots in both solar hemispheres, at sunspot minima.

* It will be observed that the median and quartiles of Figure 5 do not agree with the curves given previously (Cornish *loc. cit.*). When the original calculations were conducted, an error was made in finding the median date about 1900, which was automatically transmitted to the two quartiles and became progressively worse. This has now been rectified.

The latter phenomenon, now well known from the work of Hale and others, was first observed in 1912 (see, e.g. Chapman and Bartels 1940). Sunspots generally occur in pairs, the two spots being of opposite polarity. During any 11-year spot cycle, the polarities of the leading spots are usually the same for all pairs on the same side of the solar equator, but of opposite polarity on the other side. At the commencement of each new cycle there is a reversal of polarity, between leading and following spots in the pairs, and between northern and southern solar hemispheres. The two types of 11-year cycle, designated as *P* and *E* cycles, occurred as follows:

<i>P</i> Cycles	<i>E</i> Cycles
1843–1856	1856–1867
1867–1878	1878–1889
1889–1901	1901–1913
1913–1923	1923–1933
1933–1944	1944–

During a *P* cycle, the polarity of the leading spot of a pair in the northern hemisphere of the Sun is the same as that of the magnetic pole in the southern hemisphere of the Earth. Comparison of these intervals with those obtainable from Figure 5 shows how closely seasons having progressively early winter rains correspond with the *P* cycles, and seasons having progressively late rains correspond with the *E* cycles.

The oscillations in the Adelaide records must be due to secular changes in the latitudinal paths of anticyclones (with their attendant cyclones) across southern Australia, superimposed on the normal seasonal variation. Kidson (1925) demonstrated the pronounced seasonal variation in the latitude of the mean monthly tracks, and attempted to relate latitudinal departures from normal, at various longitudes, ranging from 120 °E. to 170 °E., to variations during the sunspot cycle. He made no differentiation of the cycles, as indicated above, so that the opposing changes noted herein have probably tended to annul each other, at 138 °E., the approximate longitude of Adelaide. In the absence of Kidson's original data, and later observations in convenient form, it is not possible to examine this point further.

The presence of a progressive change in the season at Adelaide agrees with a general observation made recently by Deacon (1953). Deacon's evidence was derived from mean differences in summer temperature, summer and winter precipitation, and barometric pressure between the two epochs 1880–1910 and 1910–1940, for stations in south-eastern Australia, and he concluded that a major climatic change has been operating progressively since 1880 (the earliest record he used), contemporaneously with changes occurring in the northern hemisphere.

The gradual advance in the latter half of the season at Adelaide also agrees with an increase in late spring and early summer rainfall at other stations as found by Mason (personal communication 1953).

IX. ACKNOWLEDGMENTS

Grateful acknowledgment is made to the Divisional Meteorologist for South Australia, who placed the Adelaide records at the author's disposal, and to Mr. M. C. Childs, of the Division of Biochemistry and General Nutrition, C.S.I.R.O., for preparation of the diagrams.

X. REFERENCES

- CHAPMAN, S., and BARTELS, J. (1940).—"Geomagnetism." (Clarendon Press : Oxford.)
CORNISH, E. A. (1936).—*Quart. J. R. Met. Soc.* **62** : 481.
DEACON, E. L. (1953).—*Aust. J. Phys.* **6** : 209.
FISHER, R. A. (1924).—*Phil. Trans.* B **213** : 89.
KIDSON, E. (1925).—*Bur. Met. Aust. Bull.* No. 17.
WHIPPLE, F. J. W. (1936).—*Quart. J. R. Met. Soc.* **62** : 492.

SHORT COMMUNICATIONS

THE CAPACITANCE OF AN ANCHOR RING*

By T. S. E. THOMAS†

The surface generated by the circumference of a circle which is rotated about a coplanar axis not intersecting the circle is usually called an anchor ring. Its capacitance in free space will depend on the radius $CP=r$ of the generating circle (Fig. 1) and the radius $OC=R$ of the centre of the generating circle. The

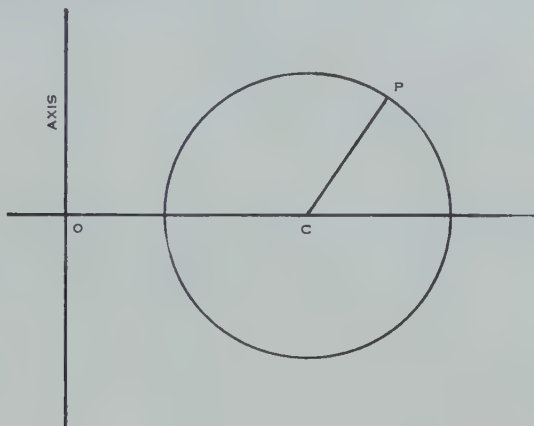


Fig. 1

solution of Laplace's equation in toroidal coordinates is discussed by Hobson (1931). Hicks (1881) has shown that the capacitance C is given by a formula involving a convergent infinite series the terms of which contain toroidal functions. The formula is

$$C=2R\left\{1-\left(\frac{r}{R}\right)^2\right\}^{\frac{1}{2}}\left[\frac{Q_0}{P_0}+2\sum_1^{\infty}\frac{Q_n}{P_n}\right], \quad \dots\dots\dots (1)$$

where

$$\left. \begin{aligned} P_n &= \int_0^{\pi} \frac{dv}{(\cosh u - \sinh u \cos v)^{n+\frac{1}{2}}}, \\ Q_n &= \int_0^{\infty} \frac{dv}{(\cosh u + \sinh u \cosh v)^{n+\frac{1}{2}}}, \end{aligned} \right\} \dots\dots\dots (2)$$

$\cosh u = R/r.$

* Manuscript received December 7, 1953.

† Dominion Physical Laboratory, New Zealand.

These integrals are the same as the Laplace and Heine integrals for Legendre functions of the first and second kind and order $n - \frac{1}{2}$. When $n=0$ and $n=1$ the integrals can be transformed into formulae involving complete elliptic integrals of the first and second kind, K and E , to modulus k where

$$k^2 = \frac{2\sqrt{R^2 - r^2}}{R + \sqrt{R^2 - r^2}}. \quad \dots\dots\dots (3)$$

K' and E' signify the same integrals to modulus k' where $k'^2 = 1 - k^2$. The formulae are :

$$\left. \begin{aligned} P_0 &= 2\sqrt{k'}K, & Q_0 &= 2\sqrt{k'}K', \\ P_1 &= \frac{2E}{\sqrt{k'}}, & Q_1 &= \frac{2(K' - E')}{\sqrt{k'}}. \end{aligned} \right\} \dots\dots\dots (4)$$

When the first two terms in the expansion have been found, the remaining terms can be found by successive applications of the recurrence formulae :

$$(2n+1)P_{n+1} - 4n \cosh u P_n + (2n-1)P_{n-1} = 0, \quad \dots\dots\dots (5)$$

$$(2n+1)Q_{n+1} - 4n \cosh u Q_n + (2n-1)Q_{n-1} = 0. \quad \dots\dots\dots (6)$$

As C is proportional to R when r/R is constant, it will be sufficient to compute C for $R=1$ and different values of r/R . The results are given in Table 1. Terms in which $Q_n/P_n < 0.0001$ were neglected. In the table C is given in e.g.s. electrostatic units, $n+1$ is the total number of terms used in the expansion, and X the sum of the terms for which $n > 1$.

When $k^2 \simeq 1$ the elliptic integrals can be obtained from the formulae

$$K = \ln(4/k')K_1 - K_2, \quad \dots\dots\dots (7)$$

$$E = \ln(4/k')E_1 + E_2, \quad \dots\dots\dots (8)$$

as tables giving the values of the constants K_1 , K_2 , E_1 , E_2 to a high degree of accuracy have been prepared by Airey (1935).

When k' is small enough $K' - E'$ is best obtained from the expansion

$$K' - E' = \frac{\pi}{2} \left[\frac{1}{2}k'^2 + \frac{3}{16}k'^4 + \frac{15}{128}k'^6 + \dots \right]. \quad \dots\dots\dots (9)$$

Integrals which could not be found from Airey's tables were obtained from Dwight's tables (1941).

If only the first two terms are of importance the formula

$$C = 2\sqrt{R^2 - r^2} \left\{ \frac{K'}{K} + 2 \frac{K' - E'}{E} \right\} = 2\sqrt{R^2 - r^2} \left\{ \frac{3K'}{K} - \frac{\pi}{EK} \right\} \quad \dots (10)$$

can be used. The error in using this is given by X in Table 1 and it will be seen that the relative error is less than 1 per cent. if $r/R < 0.46$. When only the first

term is significant, elliptic integrals can be eliminated by the approximations $K = \ln 8R/r$, $K' = \pi/2$, and $(r/R)^2 \ll 1$. The approximate formula thus obtained is

$$C = \frac{\pi R}{\ln(8R/r)} \quad \dots\dots\dots (11)$$

The relative error in using this is less than 1 per cent. if $r/R < 0.12$ and less than 4.5 per cent. if $r/R < 0.30$.

When $r/R > 0.30$ the relative error is less than 1 per cent. when the linear equation

$$C = 0.68R + 1.07r \quad \dots\dots\dots (12)$$

is used.

When C is divided by D , where $D = 2(R+r)$ is the outside diameter, it is found that C/D increases very gradually from 0.381 when $r/R = 0.30$ to 0.426 when $r/R = 0.80$, so the simple rule $C = 0.4D$ will give the capacitance with an error of less than 6 per cent. if $r/R > 0.30$. It is of interest to note that in this case the capacitance is about four-fifths that of a sphere of the same diameter.

TABLE 1

r/R	C (c.g.s., e.s.u.)	$n+1$	X
0.05	0.616	1	—
0.10	0.722	2	—
0.15	0.800	2	—
0.20	0.868	2	—
0.25	0.932	3	0.0009
0.30	0.992	3	0.0019
0.35	1.050	3	0.0037
0.40	1.106	3	0.0060
0.45	1.161	4	0.0105
0.50	1.216	4	0.0166
0.60	1.323	5	0.0409
0.70	1.429	7	0.080
0.80	1.534	7	0.163
0.90	1.638	10	0.350
1.00	1.74	—	—

In the case of a full ring with $r=R$ the Hicks formula cannot be used. However, as a full ring is the inverse surface of an infinitely long cylinder, a formula can be found by Kelvin's method of inversion (Smythe 1950).

The surface density on an earthed cylinder due to a point charge on the axis is known (Smythe 1950), and from this the surface density at corresponding points on a freely charged anchor ring can be found, together with the potential of the ring. The total charge is then found by integrating over the surface of the ring. In this way the capacitance of a full ring is found to be

$$C = 4R \sum_{s=1}^{\infty} \frac{1}{J_1(2\mu_s R)} \int_0^{\infty} \frac{\exp(-\mu_s z) dz}{\sqrt{4R^2 + z^2}},$$

where $2\mu_s R$ is given by the roots of $J_0(2\mu_s R) = 0$. On substituting $z = 2R \sinh \varphi$ this becomes

$$C = 4R \sum_{s=1}^{\infty} \frac{1}{J_1(2\mu_s R)} \int_0^{\infty} \exp(-2\mu_s R \sinh \varphi) d\varphi \quad \dots \quad (13)$$

This result is given as a problem by Smythe. As $J_1(2\mu_s R)$ is alternatively positive and negative the terms in the series will also alternate. It is obvious

that $\int_0^{\infty} \exp(-2\mu_s R \sinh \varphi) d\varphi$ is always less than $\int_0^{\infty} \exp(-2\mu_s R \varphi) d\varphi = 1/2\mu_s R$

and that the difference diminishes as s increases. The first four terms have been evaluated by numerical integration and it is found that the relative error is less than 1 per cent. when $s > 3$. The series can be summed by Euler's transformation for a slowly converging series and, using the approximate formula for the integrals when $s > 4$, the capacitance of a full ring is found to be $C = 1.74R$. This agrees with the value obtained by extrapolating the tabulated values of the Hicks formula.

References

- AIREY, J. R. (1935).—*Phil. Mag.* **19**: 177.
 DWIGHT, H. B. (1941).—"Mathematical Tables of Elementary and Some Higher Mathematical Functions." (McGraw-Hill: New York.)
 HICKS, W. M. (1881).—*Phil. Trans.* **172**: 609.
 HOBSON, E. W. (1931).—"Theory of Spherical and Ellipsoidal Harmonics." (Cambridge Univ. Press.)
 SMYTHE, W. R. (1950).—"Static and Dynamic Electricity." (McGraw-Hill: New York.)

ON THE PHOTODISINTEGRATIONS ${}^6\text{Li}(\gamma, d){}^4\text{He}$ AND ${}^6\text{Li}(\gamma, t){}^3\text{He}^\dagger$

By E. W. TITTERTON ‡ and T. A. BRINKLEY ‡

In an earlier communication (Titterton and Brinkley 1952) evidence was given of the forbidden nature of the reaction ${}^6\text{Li}(\gamma, d){}^4\text{He} - 1.54 \text{ MeV}$. Since that date Glenn (1952) has published a cross section measured at $E_\gamma = 2.76 \text{ MeV}$. Based on finding two events in a photographic emulsion loaded with the separated isotope ${}^6\text{Li}$ he gives a figure

$$\sigma_{2.76} \leq (4 \pm 4) \times 10^{-30} \text{ cm}^2, \quad \dots \quad (1)$$

which is compatible with our limit

$$\sigma_{2.76} \leq (8 \pm 2) \times 10^{-30} \text{ cm}^2. \quad \dots \quad (2)$$

† Manuscript received January 28, 1954.

‡ Research School of Physical Sciences, Australian National University, Canberra.

Our measurements at 17.6 MeV using ${}^6\text{Li}$ -loaded emulsions irradiated with γ -rays from the 440 keV ${}^7\text{Li}(p, \gamma)$ resonance have now been extended. A new lower limit for the cross section for the 17.6+14.8 MeV components of the spectrum can be given as

$$\sigma_{(17.6+14.8)} \leq (2 \pm 2) \times 10^{-30} \text{ cm}^2. \quad \dots\dots\dots (3)$$

This value is derived by assessing the irradiation in terms of the ${}^{12}\text{C}(\gamma, 3\alpha)$ events found in the emulsions and accepting the cross section for this reaction at 17.6 MeV as $2.4 \times 10^{-28} \text{ cm}^2$ (Glättli, Seippel, and Stoll 1952).

These low cross-section values are an interesting example of the isotopic spin selection rule $\Delta T = \pm 1$, not zero, for electric dipole transitions in a self-mirrored nucleus. In the experiments only $T=1$ states of the ${}^6\text{Li}$ nucleus can be formed by electric dipole transitions and these cannot break up into the two $T=0$ constituents ($\alpha + \alpha$). They could, however, break up into $T=\frac{1}{2}$ constituents, for example ($p + {}^5\text{He}$), ($n + {}^5\text{Li}$), and (${}^3\text{H} + {}^3\text{He}$) or into (${}^1\text{H} + {}^2\text{H} + {}^3\text{H}$). The first two of these reactions have been observed (Titterton and Brinkley 1951) but, to date, the reactions

$${}^6\text{Li}(\gamma, t){}^3\text{He} - 15.9 \text{ MeV}, \quad \dots\dots\dots (4)$$

and

$${}^6\text{Li} + h\nu = {}^1\text{H} + {}^2\text{H} + {}^3\text{H} - 21.4 \text{ MeV} \quad \dots\dots\dots (5)$$

have not been reported.

For this reason, in the present experiment with ${}^6\text{Li}$ -loaded emulsions irradiated with the 17.6 and 14.8 MeV γ -rays, a search has been made for the first of these reactions (4). It is energetically possible only with the 17.6 MeV component of the spectrum when it would yield events with a ${}^3\text{He}$ track of 3 μ range collinear with a triton of range 8 μ . No events have been observed and again a cross-section limit can be obtained in terms of the ${}^{12}\text{C}(\gamma, 3\alpha)$ cross section. It is

$$\sigma_{17.6} {}^6\text{Li}(\gamma, t){}^3\text{He} \leq (6 \pm 4) \times 10^{-30} \text{ cm}^2.$$

Such a low value of the cross section in the neighbourhood of the threshold is to be expected.

We wish to thank our observers, Mrs. M. Strautmanis and Misses S. Miller and L. Velioniskyte, who carried out the microscope work on this problem.

References

- GLÄTTLI, H., SEIPPEL, O., and STOLL, P. (1952).—*Helv. Phys. Acta* **25**: 491.
 GLENN, H. B. (1952).—*Phys. Rev.* **88**: 418.
 TITTERTON, E. W., and BRINKLEY, T. A. (1951).—*Proc. Phys. Soc. Lond. A* **64**: 212.
 TITTERTON, E. W., and BRINKLEY, T. A. (1952).—*Proc. Phys. Soc. Lond. A* **65**: 1052.

DELAYED DISINTEGRATION OF HEAVY FRAGMENTS*

BY FRANCOISE A. BRISBOUT† and V. D. HOPPER†

Two examples of the delayed disintegration of heavy fragments somewhat similar to those observed by Danysz and Pniewski (1953), Tidman *et al.* (1953), and Crussard and Morellet (1953)[†] have been found in this laboratory. They were observed in Ilford G5 400 μ emulsions flown at 90,000 ft.

The first event is shown in Figure 1. A heavy fragment ($Z=3$ or possibly 4) is emitted by a large star containing 2 minimum tracks, 2 grey tracks (grain density between minimum and $5 \times$ minimum) and 15 black tracks. The heavy



Fig. 1.—Event showing delayed disintegration of a heavy fragment ($Z=3$) emitted from star A and producing star B after traversing $370\ \mu$ in the emulsion.

fragment has a track length of 370 μ and reaches within 70 μ of the end of its range. If it has a charge of $Z=3$, its kinetic energy is less than 25 MeV when it produces the second star B which has two additional tracks. Of these, one short track is 15 μ long and the second track has a grain density $3 \times$ minimum. Certain identification of these tracks is not possible although scattering measurements suggest that the faster track is probably a proton of energy 90 MeV. If the other track carries the rest of the charge of the incident nucleus it will have a charge of 2 or 3.

* Manuscript received March 1, 1954.

† Physics Department, University of Melbourne.

‡ *Note added in Proof.*—Since this article was submitted, three further events of a similar type have been observed and are discussed in a recent review by Professor C. F. Powell in *Nature* **173**: 469 (1954).

The visible energy release of event B taking into account the binding energy of the emitted particles will be greater than 70 MeV. In order to conserve momentum one or more neutrons may also be emitted. If it is assumed that a single neutron balances the visible momenta and the heavy nucleus comes to rest before producing star B, the total energy will be 230 MeV. If more than one neutron is emitted, the total energy will be somewhat less but will still be greater than 100 MeV.

What appears to be a V-decay event occurs at a distance of $250\ \mu$ from star A and is shown at C. The two singly charged tracks cannot be identified but have grain densities of 43 ± 2 and 53 ± 3 . The difference in grain density seems

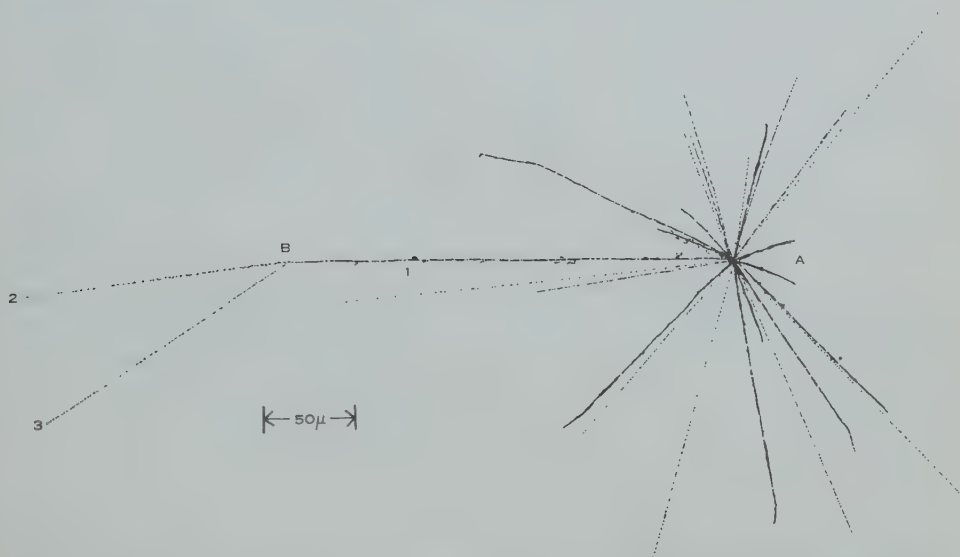


Fig. 2.—Delayed disintegration of a heavy fragment ($Z=2$) emitted from star A and producing star B before it reaches the end of its range in the emulsion.

too large to explain the event as the scattering of a single particle. The tracks are not coplanar with the origin of star A so if event C is connected with star A it cannot be completely represented by the V_2^0 or V_1^0 decay. Stars of this type are rarely observed in the nuclear emulsion but it might be merely a coincidence that it is observed close to star A. It is recorded here in case other such events are observed near stars containing heavy fragments which show delayed disintegration.

The second event is shown in Figure 2. Here the fragment ($Z=2$) is emitted from star A which has 2 minimum tracks, 3 tracks between minimum grain density and $5 \times$ minimum, and 19 other tracks. The heavy fragment track has a length of $250\ \mu$ and does not reach the end of its range before producing the second star B. If the heavy fragment is an α -particle, its energy at star B is 30 MeV. The additional tracks at star B are produced by singly charged particles, one of which must be heavier than a meson.

The total energy released in star B must therefore be greater than 150 MeV since there must be an additional uncharged particle emitted to balance momenta.

A summary of the data for the two events is given in Table 1.

TABLE 1
SUMMARY OF DATA FOR THE TWO EVENTS

Event	Track	Z	Energy (MeV)	Momentum (MeV/c)	Grain Density	Angle to Track 1
First	1	{ 3	<25	<600	Black	
		{ (4)	<(35)	<(900)		
	2	{ 2	4.3	180	Black	113°
		{ (3)	(9)	(350)		
	3	{ 1p	90	420	3.0 × min.	101.5°
		{ 1π	12	56		
Second	1	2	30	500	Black	
	2	{ 1p	90	420	3.0 × min.	8°
		{ 1π	12	56		
	3	1p	80	400	4.1 × min.	35°

The lifetimes of the heavy fragments which disintegrate have been estimated as $\geq 9 \times 10^{-12}$ and 6×10^{-12} sec for the first and second event respectively. These observations are not inconsistent with the view proposed by Danysz and Pniewski (1953) that the heavy fragments observed contain V_1^0 particles among their nucleons and these have a lifetime in the nucleus greater than 3×10^{-12} sec.

References

- CRUSSARD, J., and MORELLET, D. (1953).—*C.R. Acad. Sci., Paris* **236**: 64.
 DANYSZ, M., and PNIEWSKI, J. (1953).—*Phil. Mag.* **44**: 348.
 TIDMAN, D. A., DAVIS, G., HERZ, A. J., and TENNENT, R. M. (1953).—*Phil. Mag.* **44**: 350.

COMMENTS ON A PAPER BY E. G. BOWEN ENTITLED "THE INFLUENCE OF METEORITIC DUST ON RAINFALL". 1*

By W. C. SWINBANK†

In a paper recently published in this Journal E. G. Bowen (1953) has proposed that meteoritic dust is an important factor in stimulating rainfall. This hypothesis is advanced as a result of a study of daily rainfall statistics over a long period for Sydney and other places which reveal features, it is claimed,

* Manuscript received February 22, 1954.

† Section of Meteorological Physics, C.S.I.R.O., Melbourne.

not readily explainable in terms of known meteorological processes. It is then argued that these characteristics are attributable to dust entering the atmosphere during meteoric showers.

The speculation that some extraterrestrial influence other than solar might affect rainfall is so important that a critical review of the evidence leading to any such proposal is called for.

In effect the paper raises two issues for consideration: do the rainfall data presented contain such unexpected features as to demand special explanation and, if they do, is that offered by the author acceptable? It is proposed here to discuss the paper under these headings, and it will be contended that, for the first part, some of the evidence is inadmissible and the remainder inconclusive, and, for the second, the supposed influence of meteoritic dust is groundless.

The author begins by presenting diagrams showing the total rainfall at Sydney, day-by-day, for January and early February, for two consecutive 40-year periods. Attention is drawn to a marked "peakiness" in the distribution, in particular in respect of two dates, common to both periods, and a third which occurs only in one. In fact if the two periods are combined this last is effectively removed. Bowen states that the pattern shown in the diagrams is repeated year after year, and it is this which, in part, leads him to seek the explanation in meteoric showers. But this is not so, for the peaks are due to a few days of heavy rain and not to a greater frequency of rain on these particular dates. This point is important and will be referred to again.

Significance in the magnitude of the peaks is then implied by comparing them with the standard deviation of the remaining fluctuations from the mean. Departure from normality, however, often extreme, is commonplace in meteorological statistics, and the normal distribution is not applicable as a standard for the consideration of such an element as rainfall, for obvious reasons. A more plausible approach would be to examine the distribution of the logarithm of the rainfall. When this is done it is found that there is no significant departure from normality, and this in itself is an interesting feature.*

When reference is made to Figure 1 of the paper it becomes clear that a rainfall approaching 10 in. in one day, which is not unknown in Sydney, will dominate the statistics over the whole 40-year period and elevate that date to a major peak in the rainfall pattern. Examination of the data does, in fact, show that the peaks in the diagrams are due to heavy falls of this nature. But this feature of daily rainfall statistics has long been known and the heavy falls of rain that cause them can readily be explained by accepted dynamic processes and current theories of rain formation.

The most important link in the argument is the claim that many stations over a wide area show peaks of similar magnitude to Sydney's on "*nearly the same days*". This is the crux of the case and, if really established, would undoubtedly point to some influence on rainfall hitherto unsuspected. Bowen

* And shows that the marked troughs are as important a feature of the distribution as the peaks. They cannot be attributed to the *absence* of meteoritic dust.

states that the synchronicity of peak dates applies to many places, but presents data from eight only for January and early February in a diagram showing a grouping of peak dates, the interval between the groups being about 9 days and the spread of each group about 5–6 days. With such an interval and spread it is obvious that many stations would fit into the grouping pattern, though their peaks might be due to quite unconnected causes.

Though not stated, it is implied that the data shown for the eight stations are typical of those for the many places referred to. But the only three places for which this type of rainfall statistic has been found in the literature, Melbourne (1872–1948), Madras (1891–1940) (Ramakrishnan 1953), and Nagapattinam in south-east India (1901–1950) (Ramakrishnan and Narayanan 1953), all fail to conform.

And, if the Sydney period (1900–1949) used in Figure 3 had been extended to 1951, a fall of 224 points on January 18 in that year would have caused this date to emerge as a major peak, midway between the groups for the 13th and 23rd.

Furthermore, it is well known that as storms move they may cause rainfall intensive enough to provide major peaks at different places on successive days over an interval of several days. Clearly such events cannot fit into a pattern showing marked grouping of rainfall peaks at intervals of about 9 days.

Altogether it would seem that the most that can be said for this section is that the inference is doubtful, and that the case for synchronicity of peak dates must remain open until Figure 3 has been supplemented by the inclusion of many more stations.

In the second part of the paper the author presents evidence to show that the claimed peculiarities in rainfall at Sydney and other places, particularly the grouping of peaks, can be explained by the accession of dust to the atmosphere during meteoric showers.

Bowen states that meteoric showers have been chosen as a likely explanation because they recur year after year on the same dates. But recurrence alone would not cause the variation in rainfall pattern from one year to another, which is due to a few occasions of heavy rain. Now there is also great variation in the intensity of meteoric showers. If they exert the influence claimed it should be possible, by considering sufficient places, to demonstrate a tendency for rainfall peaks to be established in years of intense showers. The evidence shown points against this, for it is stated that rainfall peaks at different places are due to heavy falls in different years.

The relationship between meteoric showers and rainfall peaks shown in Figure 4 can, at best, only be described as slender, and almost certainly without significance. It affords no support for the assertion that peaks follow showers at Sydney after an interval of 29 days. (Note that the peak for January 31 does not appear in the 1859–1901 period.)

Consideration is next given to the physical processes by which meteoric particles might influence rain formation, and Bowen suggests that they might provide condensation nuclei, though their composition is uncertain. He quotes

a likely concentration of one particle per cubic metre, and this should be compared with the number of droplets in the average cloud, which is of order $10^8/\text{m}^3$. It is true that modern ideas on rain formation envisage the dominating influence of a relatively few large nuclei, but even so this concentration seems altogether too meagre and one cannot agree with the author that "meteoritic dust exists in adequate quantities . . .".

Having claimed an interval of 29 days between meteor showers and rainfall peaks (as evidenced by Fig. 4) Bowen then states that this period agrees with the time of fall of the meteoritic particles into the upper troposphere. Calculation shows that, for a 4μ diameter particle of assumed density 2.0, the time is in fact about 50 days. Estimates published in the literature vary from about 50 days for the larger particles to a year or two for the smaller ones. In view of these long periods, and their range, it is clearly fruitless to seek in the time of fall support for an explanation of rainfall peaks which occur at approximately 10-day intervals.

It is important to note that, in consequence of the great range in time of fall of the different sized particles compared with the interval between showers, variations in their concentration in the upper atmosphere will be very much reduced in the upper troposphere. This tendency will be strongly supported by the influence of turbulence. Variations in concentration, such as it is, at these levels are in fact more likely to be caused by atmospheric motions and to be quite unconnected with meteor showers.

This review has so far been concerned with matters directly emerging from Bowen's paper. There are other considerations which also point to the unlikelihood of meteoritic dust affecting rainfall in the manner claimed. The peaks in the rainfall pattern, it has been stated, are due to heavy individual falls which, in the case of Sydney, may exceed 10 in. and elsewhere even more. Now the total amount of precipitable water held in the atmosphere is a small number of inches, one or two say in the latitude of Sydney. In order to produce 10 in. of rain, therefore, the whole troposphere must be replaced, by vertical movement, something more than five times. For tropical places this figure could be increased to 10 or more. Even if one is prepared to accept that meteoritic dust can upset the colloidal stability of the upper troposphere, it is inconceivable that it can also influence dynamic processes on this scale. Indeed it seems likely that, if the dust is to influence rainfall at all, it would be in respect of frequency and not intensity. On the average about 2 days in 5 in Sydney are rain days and therefore over a period of 40 years there would be indication enough of any such effect. But none has been shown, and Bowen himself states that the peaks are not due to more frequent rain on those dates.

One other line of evidence can be drawn from the after effects of such enormous eruptions as Krakatoa (1883), Mont Pelé (1902), and Katmai (1912). Brilliant sunsets and unusual twilight glows on a world-wide scale long after these events, particularly in the case of Krakatoa, indicated the continued presence in the atmosphere of tiny volcanic dust particles. A measure of the size of the particles was afforded through the frequent appearance of the corona round the sun called Bishop's Ring, the angular width of which corresponded

to a particle size of about $2\ \mu$ diameter. This is comparable with the size of meteor particle quoted by Bowen, and their persistence for many months is confirmation of the times of fall for such particles given above. It is not certain that the volcanic particles were the same as the meteoritic dust, but it is known that their compositions have much in common, and one would be as likely as the other to act as condensation nuclei. But examination of the rainfall data for many stations for a considerable time after Krakatoa reveals nothing unusual, and no reference to any of these eruptions affecting rainfall has been found in the literature.

References

- BOWEN, E. G. (1953).—*Aust. J. Phys.* **6**: 490-7.
 RAMAKRISHNAN, K. P. (1953).—*Indian J. Met. Geophys.* **4**: 123-44.
 RAMAKRISHNAN, K. P., and NARAYANAN, J. (1953).—*Indian J. Met. Geophys.* **4**: 310-38.

COMMENTS ON A PAPER BY E. G. BOWEN ENTITLED "THE INFLUENCE OF METEORITIC DUST ON RAINFALL". 2*

By D. F. MARTYN†

In this Journal Bowen (1953) has recently put forward evidence purporting to show that "there is a marked tendency for heavy falls of rain to occur on certain days rather than on others, and for this pattern to be repeated year after year". It is claimed that these days of heavy rain coincide (within a margin of variation of ± 2 days) in many parts of the world, and a theory is put forward explaining the supposed phenomenon in terms of world-wide concentrations of dust produced by the impact of known meteor showers on the atmosphere. He suggests that debris from these meteors, partially disintegrated by atmospheric friction at heights of about 100 km, filters down to the tropopause in about 29 days, where it activates or "seeds" rain clouds, so producing heavy rainfall.

One purpose of this note is to show that Bowen has produced no real evidence of the existence of the world-wide phenomenon he describes; that the various rainfall peaks he selects are due to one or two *local* cyclones; and that it is possible to "establish" peaks on *any* calendar date by following his methods.

Analysis of Rainfall Data

It is notorious that of all meteorological data available for statistical analysis rainfall measurements are the least satisfactory; these can and do fluctuate

* Manuscript received March 3, 1954.

† Radio Research Board, C.S.I.R.O., Canberra, A.C.T.

wildly from day to day, and from place to (nearby) place ; one hour of torrential rain may profoundly alter the run of statistics collected over many years. Therefore, before basing conclusions upon rainfall data it is doubly essential to apply the tests of "reality" which modern statistical theory has given us : this, unfortunately, Bowen does not do.

To illustrate in a simple way the erroneous conclusions which may be drawn from inadequate treatment of rainfall data it is proposed first to examine statistics for Sydney, over the period 1902-1944, for one day, January 13. This day (Bowen 1953, Fig. 1 (b)) gives the greatest aggregate rainfall of any January day over the 43-year period : its total rainfall, summed over 43 years,

TABLE 1

SOME EXCEPTIONAL RAINFALLS IN QUEENSLAND, NEW SOUTH WALES, AND VICTORIA DURING JANUARY 1911

Place	District	Date	Rainfall for One Day (points)	Normal January Rainfall for One Day (points)
		January		
West Leichhardt	Q. (Upper Western)	8	798	11
Cloncurry	Q. (Lower Carpentaria)	9	845	14
Muckadilla	Q. (Maranoa)	10	407	11
Kenilworth	Q. (Moreton)	11	911	25
Hebel	Q. (Warrego)	12	450	5
Sydney	N.S.W. (Metropolitan)	13	708	12
Eden	N.S.W. (South Coast)	14	665	10
Walhalla	Vic. (Gippsland)		552	10

is 1140 points (11.4 in.). It is readily calculated from Bowen's diagram that the *mean* aggregate daily rainfall, averaged over all January days, is 470 points. Thus January 13 has more than double the average for January days, and is accounted a significant peak (his Table 1 and Fig. 3) derived from an excess of rain "repeated year after year" on this particular date.

Now consider the rainfall on January 13 of *one year only*, namely, 1911. On that one day, according to Hunt (1912), 708 points fell in Sydney. Deducting this one fall from the 43-year aggregate leaves only 432 points to be credited to the remaining 42 January 13ths. Thus over 42 of the 43 years this date receives *less* rain than the January daily average, though classed by Bowen as a peak rainfall day of major significance.

It seems difficult indeed to derive any worthwhile conclusion from data so variable and so limited, and impossible to attach significance to daily peaks which, contrary to the quoted statement, do *not* derive from a repeated yearly pattern, but are built of one major fall. In fact, it should be possible by such methods, and judicious choice of place and epoch, to "establish" peaks of rainfall on any chosen date. We now do this (for a limited number of days), not only

as a *reductio ad absurdum*, but because in the process light is thrown upon the true origin of such peaks. Using Hunt's (*loc. cit.*) rainfall figures for Australia during January 1911, Table 1 has been prepared. For each selected day (between January 8 and 14), a place was located which had received more than 35 times its average daily fall. It is then virtually certain that each of these places will exhibit a prominent peak on this calendar day, when the data of 43 years are

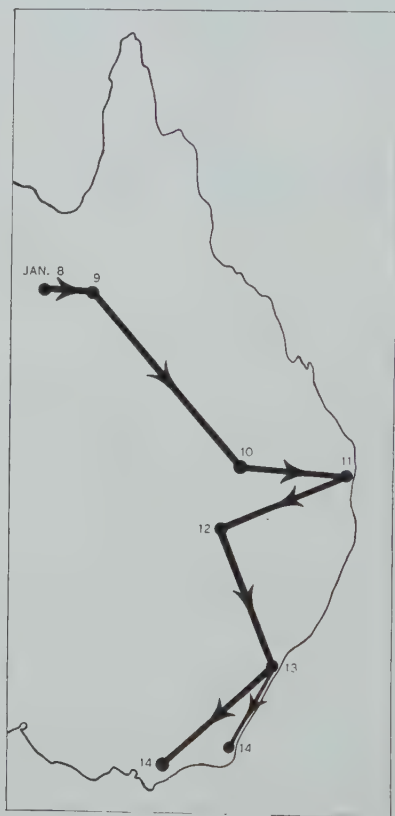


Fig. 1.—Centres in Queensland, New South Wales, and Victoria which successively have notable rainfall peaks on January 8, 9, 10, 11, 12, 13, 14 for the period 1902–1944, each the result of one heavy fall produced by the travelling cyclone of January 1911.

summed. It can be seen from the table that the rainfall figures of West Leichhardt (Upper Western Queensland), if examined by Bowen's method for 43 years, will show a peak on January 8, Cloncurry on January 9, and so on down to Eden (N.S.W.) and Walhalla (Vic.) which peak on January 14. The daily march of the sites of excessive rainfall from Carpentaria southwards is interesting (Fig. 1). As one might expect, a travelling cyclone was associated with all these heavy falls; its track (Hunt 1912) follows closely in time and place the track of Figure 1.

Clearly, daily rainfall peaks derived from data covering only 50 years merely reveal the dates on which one or two active cyclones visited the selected region. The evidence of Table 1 shows that peaks can be obtained for *any* chosen day simply by searching the rainfall figures of stations in the vicinity of a cyclone particularly active in any of the chosen years. The period January 8-14 has been here chosen for illustrative purposes since during this week the centre of the cyclone of 1911 was over the Australian mainland and could be followed from day to day. However, this cyclone produced very heavy rainfall in Australia for at least a fortnight, as the following two quotations from Hunt (1912) show :

“ *Monsoonal Depressions (of January 1911).*—The first of these influences (1st to 17th) was one of the most noteworthy monsoonal rainstorms on record. Starting in the Gulf of Carpentaria, it moved very slowly southward over the district at the head of the Gulf, thence through central, eastern, and south-eastern Queensland into the central eastern parts of New South Wales. Over a wide area on either side of its central track in Queensland and New South Wales continuous and heavy rains fell, and in Northern Queensland some phenomenal falls were recorded, e.g., Port Douglas, 1954 (for forty-eight hours to 9 a.m. 2nd) ; Granada, late Donaldson, 2780 (for forty-eight hours to 9 a.m. 9th). The average annual rainfall at Granada, it may be noted, equals 2266. The rains also extended to the Straits and very heavy falls were recorded in Gippsland.”

“ *Weather Notes—January 1911. Queensland.*—In the closing days of last year a disturbance developed off Port Darwin, and, having caused very heavy rain in the far north of Northern Territory, travelled in an eastward direction to the Gulf of Carpentaria where it lodged, with practically no change of position, until the 5th. Up to that date the influence of the approaching rainstorm co-operated with a tropical disturbance to the north-east of Cooktown, and general heavy rain fell in consequence throughout the Peninsula and Northern Coast Divisions, among the more notable registrations being 420 points at Fairview, 720 at Cairns and 1164 at Port Douglas to 9 a.m. 2nd ; 1197 points at Cairns to 9 a.m. 3rd, and 700 at Ingham and 989 at Innisfail to 9 a.m. 4th. On the 5th the monsoonal storm began to move definitely on a south-easterly course till the centre reached the vicinity of Emerald whence it bore more and more southerly, finally passing into the north-eastern regions of the neighbouring state. Rain of a widespread and generally heavy nature attended the passage of this famous disturbance throughout the entire length of Queensland, and only that part of the state lying to the westward of a line joining Camooweal, Isisford, and Hungerford escaped substantial benefit. The registrations at many of the stations over which the centre of the disturbance passed were of a quite phenomenal and, in several instances, unprecedented nature.”

It is clear that this one cyclone produced rainfall peaks in local parts of Australia on each day from January 1 to 14 in the statistics for the period 1902-1944. This finding is entirely inconsistent with Bowen's conclusion that meteoritic dust tends to produce heavy rainfall all over the world on January 12-13. His remarkable histogram (Fig. 3, loc. cit.), with three major groupings around certain dates, is not evidence of a natural phenomenon; it demonstrates, as does Figure 1 above, that for statistically short periods (say 50 years) sites may be *selected*, from the wealth of world-wide data available, which will show rainfall peaks near any particular day or group of days. It should be observed, also, that, since he deals with sites and epochs yielding 2-4 peaks per month, and allows a margin of ± 2 days for coincidence, there is an even chance that a peak in any newly chosen locality will "coincide" with a peak already found.

Local Weather Singularities

Nevertheless, although no evidence appears of *world-wide* calendar days of excessive rainfall, it should not be assumed that in *local regions* there is not a tendency for certain weather singularities to recur on or about the same dates each year. The Indian monsoons provide striking examples of such regularities: the European "summer monsoon" of early June is less marked, but equally real (Brooks 1946).

Folk-lore of the older countries contains numerous suggestions that abnormal weather tends to recur at certain times of the year. In Europe there is "Blackthorn winter" at the end of March, the "Ice saints" of May 11-13, the "Old wives' summer" of late September, "St. Luke's summer" in mid October, and the "Christmas storm". These legends were ignored by meteorologists until Buchan (1869) published a paper entitled "Interruptions in the regular rise and fall of temperature in the course of the year". Buchan believed that he had found (for Edinburgh) six short periods of cold spells and three abnormally warm periods. Buchan's work aroused great interest, and his "hot" and "cold" spells are still frequently referred to by laymen, particularly in Scotland.

In recent years Brooks (loc. cit.) has published a statistical examination of annual weather "singularities" in Western Europe, made in the British Meteorological Office in 1941-1942, obviously with a view to possible military application. In this study Brooks examined one of the most fundamental of weather phenomena, the pressure distribution. After careful statistical investigation he concludes that at least part of European folk-lore, and a part of Buchan's conclusions, are well founded. For example, he finds, from 52 years' data, a high probability that early January will be stormy (cyclonic), mid January quiet (anticyclonic). It is intrinsically likely that weather instabilities should appear regularly at this time of the year in Western Europe. To quote Brooks: "In winter the land masses, and the air over them, cool rapidly, while the surface of the ocean stays relatively warm. A pool of cold air collects over Europe, retained in place for a time by friction with the surface. Ultimately this pool becomes unstable and breaks out towards the Atlantic as a stream of cold air, after which the process begins again. Since solar radiation varies

little from one year to another, the growth of the pool of cold air to the point of instability might be expected to take about the same time in different years, so that the alternations of warm and cold periods in Britain should recur at or near the same dates."

I suggest that the "British Rainfall" data displayed in Figure 2 of Bowen's paper agree with Brooks's findings *for this region*. Bowen's data are confined to the years 1919-1949, whereas Brooks covers the period 1889-1940; both authors find a rough 10-day cycle, a stormy period early in January, and a fine spell in mid January. In view of the general agreement between Bowen's British rainfall cycle in January and the British storm cycle previously found by Brooks, there is no reason to doubt the physical reality of the former, in spite of the apparent inadequacy of the 30 years' rainfall data used. The explanation undoubtedly lies in the fact that Bowen's data, in this instance, are derived from *all* the rain-gauges in Britain; the paucity of years is substantially compensated by the abundance of observers; his "wet days" are those on which heavy rain was recorded *anywhere* in Britain.

In concluding this brief discussion of Bowen's British rainfall statistics, which, for the reason just given, appear much more soundly based than those used for the southern hemisphere, it is important to point out that these stormy periods in Britain are, as in Australia, associated with the normal travelling cyclonic disturbances which bring storms and rain to any part of the world. There can be no question of cosmic dust releasing heavy rain on any particular day or days of the year, as Bowen suggests; during their whole lives such cyclones bring storms and rain to the regions they traverse. The fact that Bowen's "wet day" cycle agrees with Brooks's cyclone cycle shows that no novel process of rainmaking is required. Energy of at least 10^{24} ergs must be supplied to create a cyclone; this vast amount (much greater than has been released in all nuclear explosions to date) cannot come from a sprinkling of dust; it comes from the solar furnace.

Noctilucent Clouds and Meteoritic Dust

Bowen concludes by demonstrating pictorially a correlation between Vestine's (1934) data on the occurrence of noctilucent clouds and Lovell and Clegg's (1952) data on meteor showers, as detected by radio means. On the limited meteor data available in 1934, Vestine had concluded that these clouds, which occur at a height of about 82 km, were probably composed of cosmic (i.e. non-terrestrial) dust. This conclusion is intrinsically likely, since meteors disintegrate near this level owing to friction with the atmosphere. Moreover, 82 km is a level of temperature inversion (Martyn and Pulley 1936) at which the atmosphere becomes dynamically stable. In the lower atmosphere the concentration of smoke and haze at such inversion levels is well known. Applying Lovell and Clegg's data Bowen has considerably improved Vestine's correlation of cosmic dust showers with noctilucent clouds. It is now virtually certain that noctilucent clouds are the visible manifestation (by reflected sunlight) of the deposition of a tenuous layer of cosmic dust at great heights. As it settles slowly through the turbulent atmosphere below 82 km this dust layer will be

rapidly dissipated and spread out over great depths. Its gradual arrival in the weather region of the troposphere must be spread over weeks or months, and could have no world-wide effect on the rainfall of any particular calendar day. But this point needs no belabouring in the absence of real evidence that world-wide rainfall singularities exist.

References

- BOWEN, E. G. (1953).—*Aust. J. Phys.* **6**: 490-97.
 BROOKS, C. E. P. (1946).—*Weather* **1**: 107-13, 130-4.
 BUCHAN, A. (1869).—*J. Scot. Met. Soc.* **2**: 4-15.
 CARRUTHERS, N. (1945).—*Quart. J. R. Met. Soc.* **71**: 144-50.
 HUNT, H. A. (1912).—*Aust. Mon. Weath. Rep.* **2**, No. 1.
 LOVELL, B., and CLEGG, J. A. (1952).—"Radio Astronomy." (Chapman and Hall Ltd.: London.)
 MARTYN, D. F., and PULLEY, O. O. (1936).—*Proc. Roy. Soc. A* **154**: 455-86.
 VESTINE, E. H. (1934).—*J. R. Astr. Soc. Can.* **28**: 249-72, 303-17.

CORRIGENDUM

VOLUME 7, NUMBER 1.

Pages 113-17, Table 2, Lists 1-5: The column headings of Columns 7 and 8 have been reversed (Error in Time of Rising and Error in Azimuth at Rising).

ON THE COULOMB AND HULTHÉN POTENTIALS

By S. T. MA*

[*Manuscript received February 26, 1954*]

Summary

This paper contains the results of two investigations. (i) The Hulthén potential is considered as a modification of the Coulomb potential. The transition from the former to the latter is investigated. (ii) The zeros of the Fredholm determinant for the Hulthén potential are calculated from the first two terms of its power-series expansion. These approximate results are compared with the exact values.

I. INTRODUCTION

It is well known that the general theory of scattering is not immediately applicable to the case of the Coulomb potential because it decreases too slowly as the distance increases (Mott and Massey 1949). In order to apply the general theory to the scattering by the Coulomb field, modifications of the Coulomb potential have been considered in the literature. A procedure introduced by Gordon (1928) consists in cutting off the field at a large distance R and making R go to infinity in the final result. Another procedure consists in replacing the Coulomb potential by the Yukawa potential, and subsequently making the exponential factor approach unity. Gordon's treatment leads to closed asymptotic expressions for large values of R . Treatment along the other line has been carried out by the method of Born approximations.

An object of the present paper is to investigate the use of the Hulthén potential (Hulthén 1942, Rosenfeld 1948) as a modified form of the Coulomb potential. Since the wave equation for the S -states can be solved exactly in the case of the Hulthén potential, we are able to trace the transition from the modified to the true Coulomb potential without making approximations.

A second issue that concerns us in this paper is the determination of discrete bound states from the zeros of the Fredholm determinant for a scattering problem. This has recently been done in work on field theories (Fubini 1953 ; Green 1954). In field theories it is often the case that the Fredholm determinant can be evaluated only in the form of a power series, and the first few terms of the series alone can be computed without excessive calculations. It is therefore of interest to see whether reasonably good results can be obtained for bound states by considering just the first few terms of the Fredholm determinant.

As a test, we consider the case of the Hulthén potential. We shall calculate the zeros of the Fredholm determinant from the first two terms of its expansion, and compare them with the zeros of its exact form. It will be seen that the

* School of Physics, University of Sydney.

approximate calculation, rough as it is, gives reasonably well the general trend of the zeros for coupling not too strong, although the numerical agreement is not good for the binding energies of the bound states.

II. TRANSITION FROM THE HULTHÉN TO THE COULOMB POTENTIAL

We write the wave equation for the S -states in the form

$$\left[\left(\frac{d}{dr} \right)^2 + k^2 \right] u(r) = U(r)u(r), \quad \dots \dots \dots (1)$$

where k^2 is the total energy E divided by $\hbar^2/8\pi^2m$ and $U(r)$ is the potential energy $V(r)$ divided by the same factor. In the case of the Hulthén potential, $U(r)$ may be written in the form

$$U(r) = -\frac{\lambda\mu e^{-\mu r}}{1 - e^{-\mu r}} \quad (\mu > 0). \quad \dots \dots \dots (2)$$

For positive-energy continuum states, k is a positive real number. The solution of (1) that has the asymptotic form e^{-ikr} is (Jost 1947)

$$f(k, r) = e^{-ikr} F(a, b, c, z), \quad \dots \dots \dots (3)$$

where F is the hypergeometric function (Magnus and Oberhettinger 1949) with

$$a = \left(\frac{ik}{\mu} \right) \left[1 - \left(1 - \frac{\lambda\mu}{k^2} \right)^{\frac{1}{2}} \right], \quad \dots \dots \dots (4)$$

$$b = \left(\frac{ik}{\mu} \right) \left[1 + \left(1 - \frac{\lambda\mu}{k^2} \right)^{\frac{1}{2}} \right], \quad \dots \dots \dots (5)$$

$$c = 1 + 2\frac{ik}{\mu}, \quad \dots \dots \dots (6)$$

and

$$z = e^{-\mu r}. \quad \dots \dots \dots (7)$$

The value of $f(k, r)$ at $r=0$ is

$$f(k, 0) = F(a, b, c, 1), \quad \dots \dots \dots (8)$$

or

$$f(k, 0) = \frac{\Gamma(c)}{\Gamma(1+a)\Gamma(1+b)}, \quad \dots \dots \dots (9)$$

or

$$f(k, 0) = \prod_{n=1}^{\infty} \frac{k + i\gamma_n}{k - in\mu/2}, \quad \dots \dots \dots (10)$$

with

$$\gamma_n = \frac{1}{2} \left(\frac{\lambda}{n} - n\mu \right). \quad (n=1, 2, \dots) \quad \dots \dots (11)$$

The matrix element $S(k)$ of the S -matrix for an S -state, which is connected with the phase $\delta(k)$ by the relation

$$S(k) = e^{2i\delta(k)}, \quad \dots \dots \dots (12)$$

is given by

$$S(k) = \frac{f(k)}{f(-k)}, \quad \dots \quad (13)$$

or, on account of (9),

$$S(k) = \frac{\Gamma(1+a^*)\Gamma(1+b^*)\Gamma(c)}{\Gamma(1+a)\Gamma(1+b)\Gamma(c^*)}. \quad \dots \quad (14)$$

For each positive real γ_n there is a bound state with the eigenvalue

$$E_n = -\frac{\hbar^2 \gamma_n^2}{8\pi^2 m}. \quad \dots \quad (15)$$

The corresponding wave function is given by (3) with $k = -i\gamma_n$,

$$a = -n, \quad \dots \quad (16)$$

$$b = \frac{\lambda}{n\mu}. \quad \dots \quad (17)$$

The first two normalized wave functions are

$$u_1(r) = \left(\frac{1}{\mu}\right) \left[\lambda \left(\frac{\lambda^2 - \mu^2}{2}\right)\right]^{\frac{1}{2}} (1-z) e^{-\gamma_1 r}, \quad \dots \quad (18)$$

$$u_2(r) = \left(\frac{1}{16\mu}\right) [\lambda(\lambda^2 - 16\mu^2)]^{\frac{1}{2}} (1-z) \left[4 - \left(2 + \frac{\lambda}{\mu}\right)(1-z)\right] e^{-\gamma_2 r}. \quad \dots \quad (19)$$

We pass from the Hulthén potential energy to the Coulomb potential energy

$$V(r) = -\frac{Ze^2}{r} \quad \dots \quad (20)$$

by putting

$$\lambda = \frac{8\pi^2 m Ze^2}{\hbar^2}, \quad \dots \quad (21)$$

and making μ approach zero. For small μ we have

$$\frac{\Gamma(1+a^*)}{\Gamma(1+a)} = S'(k) + 0(\mu), \quad \dots \quad (22)$$

where

$$S'(k) = \frac{\Gamma\left(1 - \frac{i\lambda}{2k}\right)}{\Gamma\left(1 + \frac{i\lambda}{2k}\right)} \quad \dots \quad (23)$$

is a familiar expression in the theory of Coulomb scattering, and $0(\mu)$ means a term that vanishes as μ tends to zero. Also,

$$\frac{\Gamma(1+b^*)\Gamma(c)}{\Gamma(1+b)\Gamma(c^*)} = e^{i\omega(k)}, \quad \dots \quad (24)$$

with (Jahnke and Emde 1945)

$$\omega(k) = \left(\frac{\lambda}{k}\right) \ln \left(\frac{2k}{\mu}\right) + 0(\mu). \quad \dots\dots\dots (25)$$

Thus

$$S(k) = S'(k)e^{i\omega(k)}, \quad \dots\dots\dots (26)$$

with $\omega(k)$ given by (25). This result is similar to the previous result of Gordon (Gordon 1928 ; Møller 1946), which is given by (23), (25), and (26), with $1/R$ instead of μ .

Equations (25) and (26), which are valid for small μ , may also be useful if one uses the Hulthén potential to represent a screened Coulomb potential.

In the limit $\mu=0$ we see from (11) that γ_n becomes simply $\lambda/2n$. There are then an infinite number of bound states having the energy spectrum $E_n = -\hbar^2\lambda^2/32\pi^2mn^2$ of an electron in the Coulomb field. It is easy to verify that the normalized wave functions (16) and (17) pass over to the normalized Coulomb wave functions of the $1s$ and $2s$ states respectively.

III. APPLICATION OF THE FREDHOLM THEORY TO THE HULTHÉN POTENTIAL

According to the Fredholm theory of integral equations (Whittaker and Watson 1946), the integral equation

$$X(s) = Y(s) + \lambda \int_A^B K(s,t)X(t)dt, \quad \dots\dots\dots (27)$$

under certain general conditions on the kernel K , has the solution

$$X(s) = Y(s) + D^{-1} \int_A^B D(s,t)Y(t)dt, \quad \dots\dots\dots (28)$$

if the Fredholm determinant D does not vanish. If the kernel K is such that D vanishes, there exists a solution of the homogeneous equation

$$X(s) = \lambda \int_A^B K(s,t)X(t)dt. \quad \dots\dots\dots (29)$$

The solution of (1) that describes a state of positive energy $\hbar^2k^2/8\pi^2m$ in the scattering problem satisfies an integral equation of the form (27). The kernel $K(r, r')$ satisfies the differential equation

$$\lambda \left[\left(\frac{d}{dr} \right)^2 + k^2 \right] K(r, r') = U(r) \delta(r - r'), \quad \dots\dots\dots (30)$$

and the integral term $\lambda \int_0^\infty K(r, r')u(r')dr'$ varies as $\exp(ikr)$ for large r . The Fredholm determinant D is a function of λ and k . Since we are concerned with the dependence of D on k rather than on λ , we shall denote it by $D(k)$ from now on. Corresponding to any k such that

$$D(k) = 0, \quad \dots\dots\dots (31)$$

there is a wave function $u(r)$ satisfying the homogeneous integral equation

$$u(r) = \lambda \int_0^\infty K(r, r') u(r') dr'. \quad (32)$$

There is no real k that satisfies (31). If k is an imaginary number, $i\gamma$, the wave function $u(r)$ varies as $e^{-\gamma r}$ for large r and thus describes a bound state of energy $-\hbar^2\gamma^2/8\pi^2m$ when $\gamma > 0$. There may also be complex zeros of $D(k)$ which correspond to complex energies of radioactive states.

According to the idea of analytic continuation adopted by Heisenberg in the theory of S -matrix (Heisenberg 1946; Møller 1946; Jost 1947), bound and radioactive states are connected with imaginary and complex values of k satisfying the equation

$$f(k) = 0. \quad (33)$$

That equations (31) and (33) give the same results for the problems under consideration here can be seen from the identity

$$D(k) = f(-k) \quad (34)$$

proved by Jost and Pais (1951).

In the case of the Hulthén potential, the Fredholm determinant is known in closed form on account of (34). The zeros of $f(k)$ or $D(k)$ are all purely imaginary. There are no complex zeros that correspond to complex energies.* The power-series expansion of $D(k)$ may be obtained by expanding $f(k)$ in powers of λ or from the general formula of the Fredholm theory. As already explained in Section I, it is our intention to compare the zeros of the first two terms of $D(k)$ with the exact zeros.

Expanding in powers of λ and retaining only the first two terms, we obtain from (4) and (5)

$$\Gamma(1+a) = 1 - \frac{i\lambda C}{2k}, \quad (35)$$

$$\Gamma(1+b) = \Gamma\left(1 + \frac{2ik}{\mu}\right) \left[1 - \frac{i\lambda}{2k} \Psi\left(\frac{2ik}{\mu}\right)\right], \quad (36)$$

where C is Euler's constant and Ψ is the logarithmic derivative of the factorial function (Jahnke and Emde 1945). Substituting in (9) and using (6), we obtain, up to the term linear in λ ,

$$f(k) = 1 + \frac{i\lambda}{2k} \left[\Psi\left(\frac{2ik}{\mu}\right) + C \right]. \quad (37)$$

* In this respect the Hulthén potential is different from the square-well potential or the following potential:

$$V(r) = \begin{cases} V_0 & (r_1 < r < r_2) \\ 0 & (r < r_1 \text{ or } r > r_2) \end{cases} \quad (V_0 > 0),$$

The complex energies for the last potential are known from the Gamow-Condon-Gurney theory (Møller 1946). According to Blatt and Weisskopf (1952), there exist complex energies for the square-well potential.

Let $-i\gamma'_n$ ($n=1, 2, \dots$) denote the zeros of the right hand side of (37). Let

$$x_n = \frac{2\gamma_n}{\mu}, \quad \dots \quad (38)$$

with γ_n given by (11), and let x'_n be similarly defined in terms of γ'_n . We have then

$$x_n = \frac{\lambda}{n\mu} - n, \quad \dots \quad (39)$$

$$1 - \frac{\lambda}{\mu} \frac{\Psi(x'_n) + C}{x'_n} = 0. \quad \dots \quad (40)$$

Each positive x_n corresponds to a bound state.

We give below some numerical results for the first three zeros of the Fredholm determinant calculated from (39) and (40).

(i) $\lambda = -\mu$:

$$\begin{array}{lll} x_1 = -2, & x_2 = -5/2, & x_3 = -10/3; \\ x'_1 = -1.4, & x'_2 = -2.4, & x'_3 = -3.3. \end{array}$$

(ii) $\lambda = \frac{1}{2}\mu$:

$$\begin{array}{lll} x_1 = -1/2, & x_2 = -7/4, & x_3 = -17/6; \\ x'_1 = -0.24, & x'_2 = -1.8, & x'_3 = -2.8. \end{array}$$

(iii) $\lambda = \mu$:

$$\begin{array}{lll} x_1 = 0, & x_2 = -3/2, & x_3 = -8/3; \\ x'_1 = 1, & x'_2 = -1.7, & x'_3 = -2.8. \end{array}$$

(iv) $\lambda = 2\mu$:

$$\begin{array}{lll} x_1 = 1, & x_2 = -1, & x_3 = -7/3; \\ x'_1 = 4.3, & x'_2 = -1.7, & x'_3 = -2.7. \end{array}$$

We have confined our attention to real solutions of (40). Since the exact form of $f(k)$ has no complex zeros, it is probable that its approximate form has no complex zeros either.

The general trend of the above approximate results seems to be as good as can be expected, although the numerical values for the binding energies of bound states are not accurate. The agreement becomes very bad, of course, in the case of strong coupling. For a very large value of λ/μ , (39) gives many bound states, but (40) only one.

IV. ACKNOWLEDGMENTS

The writer is grateful to Professor H. S. Green for a discussion on recent field-theoretical applications of the Fredholm theory of integral equations, and to Dr. J. M. Blatt for suggesting some improvements on the original manuscript.

V. REFERENCES

- BLATT, J. M., and WEISSKOPF, V. F. (1952).—"Theoretical Nuclear Physics." (John Wiley & Sons: New York.)
- FUBINI, S. (1953).—*Nuovo Cim.* **10**: 564.
- GORDON, W. (1928).—*Z. Phys.* **48**: 180.
- GREEN, H. S. (1954).—On the integral equations of quantized field theory. *Phys. Rev.* (in press).
- HEISENBERG, W. (1946).—*Z. Naturf.* **1**: 608.
- HULTHÉN, L. (1942).—*Ark. Mat. Astr. Fys.* **28A**, No. 5.
- JAHNKE, E., and EMDE, F. (1945).—"Tables of Functions." (Dover Publications: New York.)
- JOST, R. (1947).—*Helv. Phys. Acta* **20**: 256.
- JOST, R., and PAIS, A. (1951).—*Phys. Rev.* **82**: 840.
- MAGNUS, W., and OBERHETTINGER, F. (1949).—"Special Functions of Mathematical Physics." (Chelsea Publishing Co.: New York.)
- MØLLER, C. (1946).—*K. Danske Vidensk. Selsk., Mat.-Fys. Medd.* **22**, No. 19.
- MOTT, N. F., and MASSEY, H. S. W. (1949).—"The Theory of Atomic Collisions." (Clarendon Press: Oxford.)
- ROSENFELD, L. (1948).—"Nuclear Forces." Vol. 1. (North-Holland Publishing Co.: Amsterdam.)
- WHITTAKER, E. T., and WATSON, G. N. (1946).—"Modern Analysis." (Cambridge Univ. Press.)

NOTE ADDED IN PROOF

A. The above results calculated from (40) are encouraging, and suggest that one might attain quite good accuracy by using the expansion of $f(-i\gamma)$ up to the quadratic term, namely

$$f(-i\gamma) = 1 + \frac{\lambda}{\mu} f_1(-i\gamma) + \left(\frac{\lambda}{\mu}\right)^2 f_2(-i\gamma). \quad \dots\dots\dots (41)$$

As may be easily verified,

$$f_1(-i\gamma) = -\left[\frac{\Psi'(x) + C}{x}\right], \quad \dots\dots\dots (42)$$

$$f_2(-i\gamma) = \left[\frac{\Psi'(x) + C}{x^3}\right] + \left[\frac{\Psi'(x) + C}{2x^2}\right]^2 - \left[\frac{\Psi''(x) + \pi^2/6}{2x^2}\right], \quad \dots (43)$$

where $x = 2\gamma/\mu$ and the function Ψ' is that of Jahnke and Emde (1945). There is, however, a simpler method which has been used in Adelaide for improving the accuracy, as Professor H. S. Green kindly informed the writer. Applied to the problem considered here, the simplified method amounts to dealing with a modified form of $f(k)$, namely,

$$f_A(k) = f(k)e^{\sigma\lambda/\mu}, \quad \dots\dots\dots (44)$$

where σ is chosen such that the quadratic term in the expansion of $f_A(-i\gamma)$ vanishes, so that, to this degree of accuracy,

$$f_A(-i\gamma) = \frac{1-\lambda}{\mu} (f_1^2 - 2f_2)^{\frac{1}{2}}. \quad \dots\dots\dots (45)$$

Let $-i\gamma_n''$ denote the zeros of f_A and $x_n'' = 2\gamma_n''/\mu$. One then gets, using (42) and (43), the equation

$$1 - \frac{\lambda}{\mu} \left\{ \left[\frac{\Psi''(x_n'') + \pi^2/6}{x_n''^2} \right] - 2 \left[\frac{\Psi(x_n'') + C}{x_n''^3} \right] \right\}^{\frac{1}{2}} = 0. \quad \dots\dots\dots (46)$$

Dr. I. E. McCarthy of the University of Adelaide has obtained numerical values for the first zero x_1'' of (46). For our cases (iii) and (iv), his results are $x_1'' = -0.1$ and $x_1'' = 1.15$ respectively, which are much better than the results x_1' obtained from (40).

B. Application of the Fredholm theory to scattering problems has recently been investigated in several papers. In particular, Dr. P. Swan of the University of Melbourne has called the writer's attention to his recent work on the reduction of the Schrödinger and the second order linear integro-differential equation to Fredholm's equation. Our present paper differs from these papers in that we are mainly concerned with a bound-state problem.

ELECTRON SCREENING AND THERMONUCLEAR REACTIONS

By E. E. SALPETER*

[Manuscript received April 21, 1954]

Summary

In the interior of stars most atoms are ionized, but the electrostatic potential of a bare nucleus induces a spherically symmetric polarization of the surrounding electrons and nuclei. The effect of this screening charge cloud on the rate of thermonuclear reactions is investigated for the case of complete ionization of all atoms.

The charge distribution and potential of the screening cloud is calculated for two limiting cases where the electrostatic interaction energy between neighbouring nuclei is small or large compared with the thermal energy (weak or strong screening). The charge cloud is also investigated for intermediate strength of screening, for nuclear species which are rare and have a large charge.

Under most stellar conditions the impact parameter for a thermonuclear collision is much smaller than the radius of a screening cloud. For such cases, a simple formula is given relating the increase in the reaction rate to the potential of the screening cloud. Numerical values are presented for a few typical reactions. For conditions typical for the interior of ordinary main sequence stars the increase in the reaction rate is fairly small, usually less than a factor of two.

I. INTRODUCTION

The rates of many thermonuclear reactions, which can take place under various conditions in the interior of the Sun and stars, have been and are being calculated.† These reactions can be pictured as follows. At the high temperatures in stellar interiors all (or practically all) the atoms are ionized. Two bare nuclei (of charge Z_1 and Z_2 respectively) collide with each other with relative kinetic energy E , arising from the thermal motion of the gas. For the two nuclei to undergo a nuclear transformation they must approach to distances of the order of 10^{-13} cm (nuclear radius). As the particles approach each other they experience a Coulomb repulsion and the Coulomb barrier (electrostatic potential for a separation of the order of a nuclear radius) is very large compared with the mean thermal energy kT . An important factor in the reaction rate then is the barrier penetration factor, the probability of the nuclei approaching sufficiently closely for the nuclear forces to come into play.

The reaction rate is proportional to the following integral

$$\int_0^\infty dE [E^{\frac{1}{2}} e^{-E/kT}] P(E) \sigma_{\text{nuc.}}(E). \quad \dots \dots \dots (1)$$

* Research School of Physical Sciences, Australian National University, Canberra; on leave of absence from Cornell University, Ithaca, N.Y., U.S.A.

† See, for example, Bethe (1939), Gamow and Critchfield (1949), and Salpeter (1953).

The first term in this integrand is the Maxwell-Boltzmann distribution factor (probability of kinetic energy being E). The second factor P is the barrier penetration factor which depends very strongly on E and on $Z_1 Z_2$. The last factor $\sigma_{\text{nuc.}}$ is a purely nuclear factor which depends on the details of the interaction after barrier penetration and usually (but not always) varies fairly slowly with E .

In the usual calculations of reaction rates the barrier penetration factor $P(E)$ in (1) is evaluated by assuming the electrostatic interaction energy between the two nuclei to be purely $Z_1 Z_2 e^2 / r_{12}$, the Coulomb potential between two positive unscreened charges. But in stellar interiors the gas density ρ is high and the average distance a between a nucleus and neighbouring electrons and nuclei is small. Each nucleus, even though completely ionized, attracts neighbouring electrons and repels neighbouring nuclei and thus polarizes the surrounding gas somewhat. The nucleus is then completely screened by a spherically symmetric negative charge cloud. The radius R of this charge cloud is of the same order as the interparticle distance a or larger, depending on the ratio of Coulomb repulsion between neighbouring charges to the mean thermal energy. Hence, when two nuclei approach each other in a collision, each of them carries its screening charge cloud with it and this screening affects the interaction energy between the nuclei. We write the total interaction energy as

$$U_{\text{tot.}}(r_{12}) = Z_1 Z_2 e^2 / r_{12} + U(r_{12}). \quad \dots\dots\dots (2)$$

The main aim of this paper is to discuss the screening term $U(r_{12})$ and its effect on the barrier penetration factor P .

II. DEFINITION OF PARAMETERS

We shall now define a number of dimensionless parameters representing the ratios of various physical quantities.

(a) *Impact Parameter and Radius of Charge Cloud*

The rate of a thermonuclear reaction depends on (1), which involves an integral over E , the relative kinetic energy of the two colliding nuclei. For all reactions likely to occur in stellar interiors the integrand of (1) has a fairly sharp maximum at a particular energy $E_{\text{max.}}$. The exact value of $E_{\text{max.}}$ depends on a number of factors, including the temperature, Z_1 and Z_2 , where the relevant energy levels of the resulting compound nucleus lie, etc. But in all cases of practical interest $E_{\text{max.}}$ is very large compared with the mean thermal energy kT , in most cases larger than $10kT$.

Let r_c be the classical turning point for energy $E_{\text{max.}}$ in the collision between two nuclei of charge Z_1 and Z_2 , defined by

$$E_{\text{max.}} = \frac{Z_1 Z_2 e^2}{r_c}. \quad \dots\dots\dots (3)$$

Let r_n be the nuclear radius, that is, the distance where the nuclear attractive forces overcome the Coulomb repulsion.

Let a be a distance defined by the relation

$$4\pi a^3 \rho N_0 = 1; \quad a = \rho^{-1/3} (0.51 \times 10^{-8} \text{ cm}), \quad \dots \dots \dots (4)$$

where ρ is the gas density in g/c.c. and N_0 is Avogadro's number. The distance a is a measure of the interparticle distance, an average mass of $\frac{1}{3}$ a.m.u. being contained inside a sphere of radius a . Let R be the radius of the charge cloud surrounding a nucleus, that is, a distance beyond which an appreciable fraction of the nuclear charge is screened by the polarization charge cloud. We shall show in later sections that R is larger than a .

Let us consider a case in which the classical impact parameter r_c is very small compared with the charge cloud radius R . The nuclear radius r_n is always very much smaller than r_c . The barrier penetration factor $P(E)$ essentially depends only on the expression

$$\left[E - U(r_{12}) - \frac{Z_1 Z_2 e^2}{r_{12}} \right] \quad \dots \dots \dots (5)$$

for values of r_{12} between r_n and r_c , hardly at all on the potential for distances larger than r_c . Now $U(r_{12})$ must be a function which is small for $r_{12} \gg R$ and which approaches a constant value U_0 as r_{12} becomes small compared with R . Further the value of U_0 will be of the order of magnitude of $Z_1 Z_2 e^2 / R$. From this expression and (3) it follows that, for the case considered,

$$\frac{r_c}{R} \sim \frac{U_0}{E_{\max.}} \ll 1. \quad \dots \dots \dots (6)$$

Both in (5) and in the treatment of the nuclear factor, $U(r_{12})$ is needed only for $r_{12} \leq r_c$. If the inequality (6) is satisfied, $U(r_{12})$ can then be replaced by the potential at the origin, U_0 , which is independent of both E and r_{12} . Therefore the penetration factor P and the nuclear factor $\sigma_{\text{nuc.}}$ for energy E without screening are equal to the correct factors with screening for an energy $(E + U_0)$. The integral (1) is then replaced by

$$\int_0^\infty dE [(E + U_0)^{1/2} e^{-E/kT} e^{-U_0/kT}] P(E) \sigma_{\text{nuc.}}(E). \quad \dots \dots (1a)$$

Since U_0 is much smaller than $E_{\max.}$ (but not necessarily smaller than kT) the term $(E + U_0)^{1/2}$ can be approximated by $E^{1/2}$. The whole effect of screening in this case is then that the reaction rate with screening neglected has to be multiplied by the factor

$$e^{-U_0/kT}. \quad \dots \dots \dots (7)$$

Throughout this paper we shall consider only cases in which the inequality (6) holds. The problem is then reduced merely to evaluating U_0 for substitution into (7), without having to consider any details of the actual nuclear reactions involved. For most practical cases (6) is in fact satisfied. For the proton reaction in the solar interior (the Sun's main source of energy), for instance, $r_{\text{nuc.}} \sim 2 \times 10^{-13}$ cm, $r_c \sim 2 \times 10^{-11}$ cm, $a \sim 10^{-9}$ cm and $R \sim 3 \times 10^{-9}$ cm. The effect of screening at extremely high densities, where (6) no longer holds, is discussed by Schatzman (1948).

(b) *Degree of Ionization*

Let the parameter I_z be the ratio of the ionization potential of a K -shell electron in a hydrogen-like atom of charge Z to the mean thermal energy. We have

$$I_z = \left(\frac{Ze^2}{2a_{0z}} \right) (kT)^{-1} = \frac{(h/a_{0z})^2}{8\pi^2 m} (kT)^{-1}, \quad \dots\dots\dots (8)$$

where a_{0z} is the Bohr radius for such an atom and m is the electron mass. Throughout this paper we shall consider only cases for which the parameter I_z for all relevant values of the atomic charge is very small compared with unity. Expressing the temperature T in units of 10^6 °K, we have

$$I_z = 0.16 \frac{Z^2}{T} \ll 1. \quad \dots\dots\dots (9)$$

The problem is greatly simplified if the inequality (9) holds. All atoms are completely ionized and further we shall be able to treat the electrons by means of semi-classical approximations. In fact we shall find that our final results essentially do not depend on Planck's constant (except indirectly, the results depending slightly on the degree of degeneracy of the electrons).

For most cases of stellar interest the ionization parameter I_z is indeed much smaller than unity, with one important exception, the reactions of the carbon-nitrogen cycle in the Sun and other main sequence stars. For these reactions I_z is only slightly smaller than unity and the approximations of this paper are not very accurate. However, the effect of electron screening on these particular reactions has been discussed by Keller (1953) without assuming the inequality (9).

(c) *Strength of Screening Effect*

Let Z_1 be the larger of the two charges Z_1 and Z_2 of the two interacting nuclei and let z be the atomic charge of the main constituent of the gas (in most cases $z=Z_2$). The nature of the polarization charge cloud surrounding a nucleus Z_1 depends on whether the screening is "weak" or "strong". By "weak" screening we mean that the Coulomb interaction energy between this nucleus and the nearest few electrons and nuclei of the gas is small compared with the thermal energy kT . In this case the average positions of surrounding electrons and nuclei are displaced only very slightly from each other. The polarization charge cloud will then have a large radius R , containing many electrons and nuclei, the small difference between total negative and positive charge being Z_1 . This case will be discussed in Sections III and IV.

By "strong" screening we mean that the Coulomb interaction between the nucleus Z_1 and nearby nuclei z is large compared with kT . In this case the nucleus is surrounded in its immediate vicinity by electrons only, the nuclei z staying outside a sphere containing nearly Z_1 electrons which effectively screen the nucleus. This case is dealt with in Section V. The more difficult case of intermediate strength of screening is not treated exactly in this paper, but is discussed in Section VI.

(d) *Degree of Electron Degeneracy*

Let ξ be the average number of electrons per atomic mass unit in the stellar gas,

$$\xi = \sum_i \frac{x_i z_i}{A_i}, \quad \dots \quad (10)$$

where x_i is the fractional abundance (by mass) of nuclei of charge z_i and mass number A_i . The Fermi energy of the electron gas is defined as the thermodynamic potential at zero temperature. It is

$$E_F = \frac{\hbar^2}{8\pi^2 m} (3\pi^2 N_0 \xi \rho)^{2/3} = \left(\frac{3\pi \xi}{4} \right)^{2/3} \frac{(\hbar/a)^2}{8\pi^2 m}, \quad \dots \quad (11)$$

where a is the measure of interparticle distance defined in (4). Let D be the ratio of Fermi energy to the mean thermal energy kT . For all numerical work throughout this paper we shall express density in g/c.c. and temperature T in units of 10^6 °K. In these units

$$D = [26 \cdot 0 (\xi \rho)^{2/3} eV] (kT)^{-1} = 0 \cdot 30 (\xi \rho)^{2/3} T^{-1}. \quad \dots \quad (12)$$

The Fermi energy E_F is given by (11) only when the electrons are non-relativistic, that is, when E_F is much less than the electron restmass energy, which is the case for $\xi \rho \ll 10^6$. For $\xi \rho \gg 10^6$, equation (12) is replaced by

$$D = \frac{60 (\xi \rho)^{1/3}}{T}. \quad \dots \quad (12a)$$

In Section III we consider the special case of the electron gas being non-degenerate, $D \ll 1$. In Section IV the formulae are generalized for arbitrary values of D . All nuclei will be assumed to be completely non-degenerate, which is true for all but extremely high density ρ . The case of a degenerate nuclear gas for extremely large ρ is discussed by Schatzman (1948).

(e) *Ratio of Abundances and Charges*

The formulae derived in this paper for the two limits of very weak and very strong screening, hold quite generally for all values of Z_1 and Z_2 and any composition of the gas. But for many (although by no means all) cases of stellar interest, one of the two reacting nuclei Z_1 has a very low abundance x_1 and a large charge compared with that of the other reacting nucleus and with the average charge z of the gas nuclei.

For this special case,

$$Z_2, z \ll Z_1; x_1 \ll 1, \quad \dots \quad (13)$$

the problem of electron screening is simplified and numerical solutions are also presented for intermediate strength of screening in Section VI. The two earlier papers by Schatzman (1948) and Keller (1953) are mainly concerned with cases for which (13) holds.

III. WEAK SCREENING : ELECTRONS NON-DEGENERATE

We shall now consider the case of the screening being "weak", in the sense discussed in Section II (c). For simplicity we also assume in this section that the electrons are non-degenerate and consider the main constituent of the gas to be atoms of atomic charge z and atomic weight A . Our aim is to calculate the interaction energy $U(r)$, defined in (2), between nuclei of charges Z_1 and Z_2 at a distance r .

We now assume that $U_{\text{tot.}}(r)$ is of form

$$U_{\text{tot.}}(r) = Z_1 Z_2 e^2 \psi_{\text{tot.}}(r) ; \quad \psi_{\text{tot.}}(r) = r^{-1} + \psi(r), \quad \dots \quad (14)$$

where $\psi(r)$, the part due to the screening cloud, is independent of both Z_1 and Z_2 (but does depend on z). We shall show later that this assumption is justified if the screening is weak. It should be remembered that $U(r)$ represents a statistical average of the interaction energy. Only the two nuclei Z_1 and Z_2 are at a fixed separation r , the position of all other nuclei being averaged over.

Consider the nucleus Z_1 fixed at the origin and let $V(r)$, $\bar{\rho}(r)$ be the electrostatic potential and electric charge density at a point r , averaged over all particles except the nucleus at the origin. If $U(r)$ is of the form of (14), depending on the charges only through the factor $Z_1 Z_2$, then the interaction energy of an infinitesimal "test charge" δZ at the point r is $\delta Z Z_1 e^2 \psi(r)$ and hence $V(r)$ is $Z_1 e \psi(r)$. Since V and $\bar{\rho}$ are related by Poisson's equation, we have

$$\nabla^2 [Z_1 e \psi_{\text{tot.}}(r)] = -4\pi \bar{\rho}(r) - 4\pi Z_1 e \delta^{(3)}(r). \quad \dots \quad (15)$$

Another relation between ψ and $\bar{\rho}$ is furnished by statistical mechanics, namely, that the density of particles (nuclei z or electrons) in a region in which each particle has potential energy $U_e(r)$ is the field-free density times the Boltzmann factor $\exp[-U_e(r)/kT]$. Since the potential energy of electrons has opposite sign to that of nuclei, their electric charge densities no longer cancel each other exactly and we have

$$\bar{\rho}(r) = \left(\frac{\rho N_0 z e}{A} \right) \left\{ \exp \left[\frac{-Z_1 z e^2 \psi_{\text{tot.}}(r)}{kT} \right] - \exp \left[\frac{+Z_1 e^2 \psi_{\text{tot.}}(r)}{kT} \right] \right\} \dots \quad (16)$$

Substituting the expression (16) on the right-hand side of equation (15) gives an inhomogeneous second order differential equation in $\psi(r)$, the Poisson-Boltzmann equation.

According to our assumption of weak screening, the exponents of the two exponential terms in (16) are small compared with unity for r of the order of magnitude of a or larger. We can then expand these two exponential factors. The first terms in the two factors cancel each other and we keep only the second term of each, linear in Z_1 . Substituting this approximation into equation (15) we get the linear approximation to the Poisson-Boltzmann equation,

$$\nabla^2 \psi(r) = 4\pi \rho N_0 \left(\frac{z^2 + z}{A} \right) \frac{e^2}{kT} \left[\frac{1}{r} + \psi(r) \right]. \quad \dots \quad (17)$$

The boundary conditions for (17) are imposed by the physical condition that $\psi(r)$ represents the potential due to the screening charge distribution, which

has a finite radius and a total charge equal and opposite to that of the nucleus at the origin. Hence $\psi(r)$ approaches $-r^{-1}$ as r approaches infinity and approaches a finite limit $\psi(0)$ as r approaches zero.

We define the "radius" of the charge cloud, R , by

$$R^2 = \left(\frac{kT}{4\pi\rho N_0 e^2} \right) \left(\frac{A}{z^2 + z} \right) = \left(\frac{A}{z^2 + z} \right) \left(\frac{kT}{e^2/a} \right) a^2. \quad \dots\dots (18)$$

The solution of (17) is then

$$\psi(r) = r^{-1}(e^{-r/R} - 1); \quad \psi(0) = -R^{-1}. \quad \dots\dots\dots (19)$$

Our approximate equation and its solution, (17) and (19) respectively, are mathematically equivalent to those derived by Debye and Hückel (1923) in their theory for dilute solutions of electrolytes. A number of assumptions and approximations are involved in our derivation of (17) above. We shall now discuss the validity of these approximations.*

(1) We have used a continuous (average) charge density $\bar{\rho}(r)$ and in its evaluation have used the statistical Boltzmann factor $\exp[-U(r)/kT]$ for particles at the point r . For this procedure to be strictly valid many nuclei and electrons should be contained in a volume small enough so that $\bar{\rho}(r)$ and $U(r)$ do not vary appreciably over this volume. Now the linear dimension of the charge distribution is of order R , the average distance between particles of order a . Hence our procedure is valid as long as $a \ll R$. For the case of weak screening which we are considering $z^2 e^2/a \ll kT$. It then follows from (18) that R is indeed large compared with a .

(2) In our whole derivation we have treated the electrons classically and considered their charge density at a definite distance r from the origin. This assumes that we can specify the position of an electron to within a distance r without considering the corresponding Heisenberg uncertainty energy. The neglect of this uncertainty energy is justified as long as it is small compared with the mean thermal energy of the electrons. Hence our classical treatment is justified only for values of r larger than a critical distance $r_{\min.}$, which is roughly given by

$$h^2/8\pi^2 m r_{\min.}^2 \sim kT. \quad \dots\dots\dots (20)$$

We are at present only considering cases where the ionization potential for an atom Z_1 and the Fermi energy of the electron gas are both small compared with kT . A comparison of (8) and (11) with (20) then shows that the limiting distance $r_{\min.}$ is much smaller than the Bohr orbit a_0 , and also much smaller than a , which in turn is smaller than R .

(3) As was discussed above, we have expanded the two exponential factors in (16), keeping only the first two terms in the expansion. Since $\psi_{\text{tot.}}$ is of order r^{-1} for small distances, this linear approximation is valid for r larger than r_e , where $(Z_1 z e^2/r_e) \sim kT$ for nuclei z . For the case of weak screening r_e is small

* See also the discussions by Fowler and Guggenheim (1949) and by Keller and Meyerott (1952).

compared with R , the distance from which most of the screening effects stem. For the electrons one finds, using (8) and (20), that r_e is not only smaller than R , but also smaller than r_{\min} .

(4) In deriving the Poisson-Boltzmann equation we have assumed that the interaction energy between two particles has the form of (14), with $\psi(r)$ independent of Z_1 and Z_2 . The solution (19) for $\psi(r)$, obtained by means of the linear approximation, satisfies this requirement, R depending only on z , but not on Z_1 and Z_2 .

Our final answer for the screening potential between nuclei Z_1 and Z_2 is then

$$-\frac{U_0}{kT} = \frac{(Z_1 Z_2 e^2)}{R} (kT)^{-1} = 0.188 Z_1 Z_2 \zeta \bar{\rho}^{\frac{1}{2}} T^{-3/2}, \quad \dots \quad (21)$$

where

$$\zeta = (z^2 + z)^{\frac{1}{2}} A^{-\frac{1}{2}}.$$

The conditions of validity for (21) are that (9) and (12) be satisfied and that the screening is weak. More exactly, the weak screening condition requires the interaction energy between nuclei Z_1 and z at separation R to be smaller than kT ; that is, the expression (21), with Z_2 replaced by z , must be smaller than unity.

IV. WEAK SCREENING: GENERAL CASE

We first introduce a trivial generalization of the work of the preceding section. We consider the gas as made up of different nuclear species of charge z_i and mass number A_i , their fractional abundance (by mass) being x_i . The average number of electrons per a.m.u. is then the quantity ξ , defined in (10) (instead of the z/A of Section III). The first exponential factor in (16) is then replaced by a sum of exponentials, one for each nuclear species. Expanding these exponentials as before, the first terms again cancel the first term from the electron factor. The second terms are quadratic in z_i and finally in (17) the expression z^2/A is simply replaced by $\sum_i x_i z_i^2/A_i$.

The effect of electron degeneracy is slightly more complicated. We consider now the case of arbitrary degree of degeneracy, but weak screening. It should be noted that it is still possible to have a degenerate electron gas but weak screening if, and only if, the inequality (9) is satisfied. We again want to consider the electron charge density $\bar{\rho}_e(r)$ at a point r and to neglect the corresponding uncertainty energy which is roughly $(\hbar/2\pi r)^2/2m$. This neglect is justified if the uncertainty energy is much smaller than the average electron kinetic energy, which is of order $E_F \sim (\hbar/2\pi a)^2/2m$, if the electrons are degenerate. This is the case if r is larger than a . We again will be most interested in distances of the order of R , again large compared with a , which justifies our procedure.

Let $f(\eta)$ be the Fermi-Dirac function,

$$f(\eta) = \int_0^\infty dx \, x^{\frac{1}{2}} [e^{(x-\eta)} + 1]^{-1}, \quad \dots \quad (22)$$

and D be the ratio of the Fermi energy E_F to kT . For the field-free case the thermodynamic potential $\mu = \eta kT$ of the electron gas* is determined by the equation

$$f(\eta) = \frac{2}{3} D^{3/2}. \quad (23)$$

If electrons near the point r experience an interaction energy $U_e(r)$, the electron density will no longer be uniform throughout space but will adjust itself such that the total thermodynamic potential is uniform. The ratio of the actual density to the field-free density is then $f(\eta - U_e(r)/kT)/f(\eta)$, where η is still given by (23). If the screening is weak and if $a < r \sim R$, then $U_e(r) \ll kT$. We can then use a Taylor expansion for the expression $f(\eta - U_e/kT)$. Retaining only the first two terms in this expansion, the ratio of actual density to field-free density becomes

$$1 - \left[\frac{U_e(r)}{kT} \right] \left[\frac{f'(\eta)}{f(\eta)} \right], \quad (24)$$

where f' is the first derivative of $f(\eta)$. This expression is our generalization of the linear approximation $(1 - U_e/kT)$ to the Boltzmann factor, which we used in Section III.

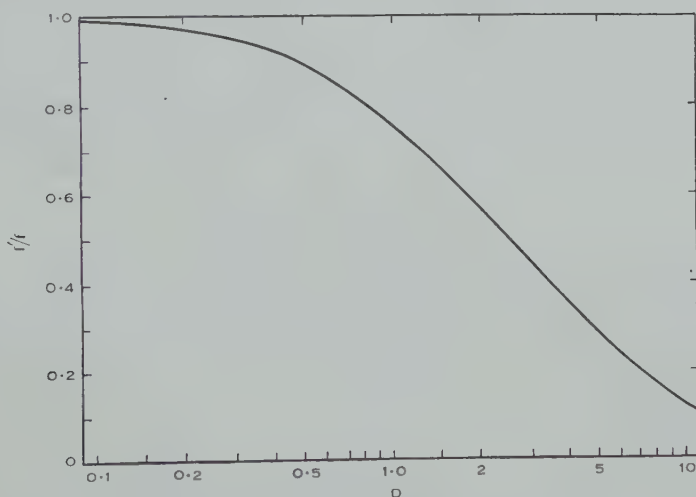


Fig. 1.—The quantity $f'(\eta)/f(\eta)$, to be substituted into equation (24), plotted against the degeneracy parameter D .

The factor f'/f takes on simple forms in two limiting cases. For extreme non-degeneracy, $D \ll 1$ and η is negative and very large. In this case f'/f becomes equal to unity and the expression (24) reduces to the expression of Section III. For extreme degeneracy, $D \gg 1$, η approaches D and f'/f approaches $(3/2D)$. In this case the polarization of the electron gas is negligible compared with that of the nuclear gas. For intermediate values of D , η and then f'/f were obtained from (22) and (23), using the numerical tabulation of the Fermi-Dirac function given by McDougall and Stoner (1938). The results of these numerical calculations are given in Figure 1, the quantity $f'(\eta)/f(\eta)$ being plotted against D .

* See, for instance, Mayer and Mayer (1940).

Our final formula for U_0/kT is then still (21), but with the quantity ζ now having the general form

$$\zeta = \left\{ \sum_i x_i \frac{z_i^2}{A_i} + \left(\frac{f'}{f} \right) \sum_i x_i \frac{z_i}{A_i} \right\}^{\frac{1}{2}}. \quad \dots\dots\dots (25)$$

V. STRONG SCREENING

We shall now investigate the screening charge cloud surrounding a nucleus Z_1 imbedded in an ionized gas in the limit of strong screening. We shall see later that the radius R of the charge cloud in this case is approximately $(3Z_1/\xi)^{1/3}a$. The condition for strong screening is then that the Coulomb repulsion between charges Z_1 and z (average charge of gas nuclei) at distance R be large compared with kT . For the sake of simplicity we restrict ourselves slightly more to the case where this Coulomb energy is also considerably larger than the ionization potential for charge Z_1 . We shall show that the electron density is then essentially uniform, which simplifies the problem greatly.

We thus assume the following conditions to apply :

$$2zI_z \ll \frac{Z_1 z e^2}{a} (kT)^{-1} \gg \left(\frac{3Z_1}{\xi} \right)^{1/3}, \quad \dots\dots\dots (26)$$

where I_z and ξ are defined in (8) and (10) respectively. From the first part of this inequality and (8) it follows that $a \ll a_{0z}$ and that

$$\frac{Z_1 e^2}{a} \ll \left(\frac{h}{2\pi a} \right)^2 m^{-1} \sim E_F.$$

Hence the Coulomb energy of an electron at distance a or greater is small compared with the Fermi energy and at distances less than a small compared with the corresponding uncertainty energy. It then follows from the discussion of Section IV that the electron density at all relevant distances from the origin differs only slightly from the field-free density. We shall therefore take the electron density to be uniform.

The second inequality in (26) expresses the condition that the Coulomb repulsion experienced by nuclei z at distances between a and R is much larger than kT , unless a very large fraction of the charge Z_1 is screened by the electron cloud. For these distances then the density of gas nuclei is very small compared with the field-free density. We can therefore use the following approximate picture for the screening charge cloud. The nucleus Z_1 is surrounded by electrons of uniform density up to a sphere of radius R_1 , where

$$R_1 = \left(\frac{3Z_1}{\xi} \right)^{1/3} a. \quad \dots\dots\dots (27)$$

No nuclei z are present within this sphere and the total charge of the electrons inside the sphere is $-Z_1$. Outside this sphere the nucleus Z_1 is effectively screened, the density of nuclei z also has its field-free value and there is no net charge density outside the sphere.

On this simple picture the total electrostatic energy of the system of nucleus Z_1 and its surrounding electron sphere is

$$W_1 = -\left(\frac{3}{2} - \frac{3}{5}\right) \frac{Z_1^2 e^2}{R_1} = -\frac{9 \xi^{1/3} Z_1^{5/3} e^2}{3^{1/3} \times 10a}, \dots\dots\dots (28)$$

where the first term is the interaction energy between the nucleus and the electron cloud and the second term is the electrostatic self-energy of the electron cloud. When two nuclei, Z_1 and Z_2 respectively, approach each other in a collision, each of them carries its own electron sphere with it until the two spheres interpenetrate. When interpenetration occurs the gas nuclei rearrange themselves, the density of electrons still remaining uniform. For intermediate distances of approach the shape of the electron cloud is rather complicated. But for distances between the nuclei small compared with R_1 and R_2 , the electron cloud is again spherical and of the same radius as the electron cloud for a nucleus of charge $(Z_1 + Z_2)$.

As was discussed in Section II, we are interested only in the interaction energy between the two nuclei and their charge clouds for separations less than a and therefore less than R_1 and R_2 . On our present approximation the screening contribution to this energy is simply the interaction energy (28), for a nucleus of charge $(Z_1 + Z_2)$, minus the sum of the equivalent energies for the two nuclei Z_1 and Z_2 . This screening potential is

$$\begin{aligned} -\frac{U_0}{kT} &= \frac{9e^2}{10akT} \left(\frac{\xi}{3}\right)^{1/3} [(Z_1 + Z_2)^{5/3} - Z_1^{5/3} - Z_2^{5/3}] \\ &= 0.205 [(Z_1 + Z_2)^{5/3} - Z_1^{5/3} - Z_2^{5/3}] (\xi \rho)^{1/3} T^{-1}. \quad \dots (29) \end{aligned}$$

Numerically, the condition (26) can be written as

$$Z_1 \ll \rho^{1/3}; \quad 0.23 Z_1^{2/3} z(\xi \rho)^{1/3} T^{-1} \gg 1. \quad \dots\dots (26a)$$

Besides assuming this inequality we have made some further approximations.

(1) As in previous sections we have used continuous charge densities. Since the number of particles making up the screening charge cloud is now only of order Z_1 , the fluctuations omitted in this treatment may not be negligible.

(2) In reality the electron density is not completely uniform, but slightly larger near the central nucleus. Similarly the density of gas nuclei does not jump discontinuously from zero inside the radius R to the field-free value just outside, but changes gradually over a certain "skin-depth". But the more strongly the two inequalities in (26) are satisfied, the more nearly uniform is the electron density and the smaller is the ratio of skin-depth to radius R .

(3) In the strong screening case the screening cloud surrounding a charge $(Z_1 + Z_2)$ is no longer a linear superposition of the clouds for charges Z_1 and Z_2 , the radii of the clouds being different. We have neglected all dynamic effects ensuing from the rearrangement of the gas nuclei, which may in fact not be negligible. (29) may therefore not be a very accurate approximation.

VI. NUCLEI OF LOW ABUNDANCE AND LARGE CHARGE

We have so far considered the general case of arbitrary abundance and charge of the nuclei Z_1 . For this general case we have seen that the use of the Poisson-Boltzmann equation is not self-consistent for intermediate and strong screening, largely due to the non-linearity of its solution. For the case of strong screening we derived an approximation independent of the Poisson-Boltzmann equation, but its accuracy is limited by the non-linearity of the screening potential and by the small number of particles involved in the charge clouds.

As was pointed out by Keller and Meyerott (1952), these difficulties are removed if the abundance of the nuclear species of charge Z_1 is low and if Z_1 is much larger than Z_2 and z (equation (13)). Since $Z_1 \gg Z_2$ we need only consider the interaction of the nucleus Z_2 with the charge cloud surrounding the nucleus Z_1 , without explicitly considering the charge cloud around Z_2 . In this approximation the radius R_1 of the charge cloud does not alter during a nuclear collision, even if the screening is strong. We are then justified in neglecting the dynamic effects of the rearrangement of gas nuclei, mentioned in Section V. Since the abundance of the nuclei Z_1 is low, their average distance from each other is large compared with the radius of their charge cloud. We can then consider the charge cloud around a single such nucleus. Since $Z_1 \gg z$, we can also neglect the charge cloud around all nuclei z . In this case the condition of linearity (14) is no longer required for the validity of the Poisson-Boltzmann equation. Finally, the charge cloud contains at least Z_1 electrons, even for strong screening. Since Z_1 is large, our neglect of fluctuations from average distributions is justified.

Keller (1953) has described numerical methods for solving the Poisson-Boltzmann equation for arbitrary screening strength, if the inequality (13) holds. For the sake of simplicity we shall consider here only cases satisfying two further conditions: (1) the bulk of the gas consists of only one type of nuclear species of charge z and mass A , (2) the electrons are degenerate or $z \gg 1$. In either case the departure from uniform density is much smaller for the electrons than for the nuclei z . We thus take the electron density to have its field-free value.

Writing the electrostatic potential at distance r from the nucleus Z_1 as $Z_1 e/r + V(r)$, the Poisson-Boltzmann equation becomes

$$\nabla^2 V(r) = 4\pi z e \frac{\rho N_0}{A} \left\{ 1 - \exp \left[-\frac{Z_1 z e^2}{r k T} - \frac{z e}{k T} V(r) \right] \right\}. \quad \dots (30)$$

The exponential term in parentheses arises from the Boltzmann factor for the nuclei, the constant term from the assumed uniform electron density. Making the substitutions

$$Y(r) = \frac{z e}{k T} V(r), \quad R^2 = \frac{A k T}{4\pi \rho N_0 z^2 e^2}, \quad x = \frac{r}{R}, \quad F = \frac{Z_1 z e^2}{R k T},$$

equation (30) takes the form

$$\nabla_x^2 Y(x) = 1 - \exp \left[-\frac{F}{x} - Y(x) \right]. \quad \dots (31)$$

The boundary conditions for (31) are that $Y(x)$ approaches a finite value at the origin and approaches $-F/x$ as x tends to infinity.

Following the discussion at the beginning of this section, we assume that the interaction energy U_0 for a collision between nuclei Z_1 and Z_2 (to be substituted into (7)) is simply $Z_2 e V(0)$. Hence

$$\frac{U_0}{kT} = \frac{Z_2}{z} Y(0). \quad \dots\dots\dots (32)$$

$Y(0)$ is the value at the origin of $Y(x)$, obtained by solving (31) for the appropriate value of the parameter F . Numerically,

$$F = 0.188 \frac{Z_1 z^2 \rho^{1/2}}{A^{1/2} T^{3/2}}. \quad \dots\dots\dots (33)$$

Equation (31) has simple analytic solutions in two limiting cases. If $F \ll 1$, it is

$$Y(x) = \frac{F}{x} (e^{-x} - 1); \quad Y(0) = -F. \quad \dots\dots\dots (34)$$

This solution is identical with our approximation for weak screening, derived in Sections III and IV. The factor ζ of (21) and (25) is here $z/A^{1/2}$, appropriate if the electrons are degenerate or if $z \gg 1$. If $F \gg 1$, the solution becomes

$$\left. \begin{aligned} Y(x) &= -\frac{3^{2/3}}{2} F^{2/3} + \frac{x^2}{6}, & \text{for } x < (3F)^{1/3}, \\ Y(x) &= -\frac{F}{x}, & \text{for } x > (3F)^{1/3}. \end{aligned} \right\} \quad \dots\dots (35)$$

This solution is identical with our approximation for strong screening, derived in Section V. The quantity U_0 is then given by

$$-\frac{U_0}{kT} = \frac{3^{2/3}}{2} \frac{Z_2}{z} F^{2/3} = 0.34 \frac{Z_1^{2/3} Z_2 z \rho^{1/3}}{A^{1/3} T}. \quad \dots\dots\dots (36)$$

If we expand the expression in parentheses in (29) in powers of (Z_2/Z_1) and retain only the lowest order term, (29) reduces to (36).

Equation (31) was solved numerically for a few intermediate values of the parameter F . It was found convenient to transform this equation into one involving $y(x) = xY(x)$. It is

$$\frac{d^2 y}{dx^2} = x \left\{ 1 - \exp \left[-\frac{(F+y)}{x} \right] \right\}, \quad \dots\dots\dots (37)$$

with the boundary conditions $y(0) = 0$ and y approaching $-F$ as x tends to infinity. The expression wanted is $Y(0)$, the first derivative of $y(x)$ at the origin. For each value of F separately, $Y(0)$ was found by trial and error, the equation being integrated from the origin outwards with different guesses for $Y(0)$. This procedure was continued until y for large x approached $-F$ for a particular value of $Y(0)$.

A graph of $Y(0)$ *v.* F is given in Figure 2. A few numerical results follow : for $F=1.5$, the numerical result for $Y(0)$ is about 0.98, as compared with 1.5 on the weak screening approximation and 1.36 on the strong screening approximation ; for $F=0.25$, $Y(0)$ is 0.21, as compared with 0.25 on the weak screening approximation ; for $F=15$, $Y(0)$ is 5.7 as compared with 6.3 on the strong screening approximation.

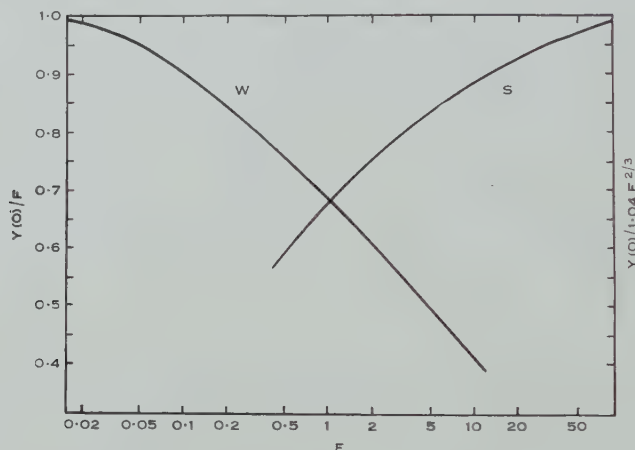


Fig. 2.—The ratio of the exact value of $Y(0)$ to its weak screening and strong screening approximations, as a function of F . The curve W gives $Y(0)/F$, the curve S gives $Y(0)/1.04F^{2/3}$.

VII. SOME NUMERICAL EXAMPLES

We conclude this paper with a few specific examples of nuclear reactions, which are of importance in different types of stars. According to (7), the effect of electron screening consists solely in multiplying the reaction rate by a factor $\exp(-U_0/kT)$. We give below numerical values for the exponent $(-U_0/kT)$.

(a) Proton-Proton Chain

This reaction chain provides the main energy source for our Sun and for all cooler main sequence stars. The reaction determining the rate of energy production involves the collision between two protons. The rate of this reaction, without any screening correction, is the most accurately known one of all thermonuclear reactions of stellar interest. It is therefore important to calculate the effect of screening fairly accurately. Fortunately the screening for this reaction in the interior of main sequence stars is quite weak and the ionization potential of hydrogen very much less than kT . Equation (21) should therefore be a good approximation.

(i) Approximate central conditions in the Sun are $\rho=100$, $T=13$. From (8) the ionization parameter I_z is about 0.01, which certainly satisfies the inequality (9). The solar interior consists mainly of hydrogen, with the abundance of helium, x_{He} , of order 0.2 or less. From (12) the degeneracy parameter D is about 0.5. From Figure 1 the factor f'/f is about 0.90. Hence

ζ , defined by (25), is about $1.38(1-0.12x_{\text{He}})$. Finally, using this value of ζ and putting $Z_1=Z_2=1$ for the proton-proton reaction, (21) gives

$$-\frac{U_0}{kT}=0.055(1-0.12x_{\text{He}}). \quad \dots\dots\dots (38)$$

(ii) For a typical red dwarf star (a main sequence star cooler than the Sun) approximate central conditions are $\rho=100$, $T=8$. In this case $D=0.8$, $f'/f=0.8$, and $\zeta=1.34$. Equation (21) then gives a value of about 0.11 for $-U_0/kT$.

(b) Carbon-Nitrogen Cycle

This chain of reactions provides the energy source for the hotter main sequence stars. The reaction determining the rate of energy production is one involving the collision between a nitrogen nucleus ($Z_1=7$) and a proton.

(i) We first consider a case for which an accurate numerical computation has been carried out by Keller* (1953), namely for $\rho=122$ and $T=11.6$. In this case the ionization parameter $I_z=0.68$. Since this value is not very small compared with unity we should not expect the formulae of the present paper to be very accurate. With the gas mainly consisting of hydrogen, $D=0.64$ and $f'/f=0.85$. Using the weak screening formula (21) we get $-U_0/kT=0.50$. Keller's accurate value for this quantity is 0.43.

(ii) The actual conditions in the deep interior of a star which derives its energy from the carbon cycle involve lower densities and higher temperature than in the above example. Hence I_z is somewhat smaller and the screening slightly weaker and (21) should be a reasonably good approximation. Conditions typical of the interior of Sirius, say, are $\rho=80$ and $T=20$. In this case I_z is about 0.34, D is about 0.3, and f'/f almost unity. (21) then gives $-U_0/kT=0.19$.

We can conclude from these few examples that the effect of electron screening on the rate of energy production in all main sequence stars is fairly small, less than a factor of two in most cases.

(c) Formation of ${}^8\text{Be}$

The more luminous main sequence stars convert hydrogen into helium at such a rapid rate that their central regions may be completely without hydrogen after a lifespan much less than the age of our Galaxy. The subsequent life history of such a star is not yet completely understood but it is likely that the core of such a star, consisting almost entirely of helium, will contract and both density and temperature will increase. When a temperature of 1 to 2×10^8 °K is reached, helium begins to be transformed into carbon and heavier nuclei. The first step in these processes is the formation of the short-lived ${}^8\text{Be}$ nucleus from the collision of two α -particles. The heavier nuclei are then built up by successive radiative captures of α -particles, starting from ${}^8\text{Be}$ (α, γ) ${}^{12}\text{C}$.

We give now values of the quantity U_0 , the screening energy, for the collision of two helium nuclei in a gas consisting largely of ${}^4\text{He}$. The order of magnitude

* See also Schatzman (1954).

of densities in stellar interiors at very high temperature is not yet known, so we consider a few different values.

(i) $\rho=10^4$, $T=150$: The ionization parameter I_z is negligibly small, $\xi=z/A$ is $\frac{1}{2}$, the degeneracy parameter $D=0.58$, and f'/f is about 0.87. The parameter ζ is then about 1.20 and (21) gives a value of 0.048 for $-U_0/kT$.

(ii) $\rho=10^6$, $T=150$: The electrons are highly degenerate ($D>10$) and f'/f is practically zero, ζ practically unity. Equation (21) gives a value of 0.40 for $-U_0/kT$, but, since this number is not very small, our weak screening approximation is probably not very accurate.

(iii) $\rho=10^8$, $T=150$: The electrons are highly degenerate (relativistically) and their density is practically uniform. The screening is fairly (but not very) strong. Hence (29) should give an approximation which is fairly poor, but at least better than (21). The result is $-U_0/kT=1.9$.

(d) Formation of ^{20}Ne

We take as our final example the collision between an oxygen nucleus and an α -particle. The reaction $^{16}\text{O}(\alpha, \gamma)^{20}\text{Ne}$ is one in the reaction chain beginning with the formation of ^8Be . We again consider a gas consisting mainly of helium at a temperature of 1.5×10^8 °K.

(i) $\rho=10^4$, $T=150$: In this case the screening is still reasonably weak. The calculation proceeds as for the collision between two α -particles (Example (a) (i)) with Z_1 being 8 instead of 2. This gives a value of 0.19 for $-U_0/kT$.

(ii) $\rho=10^6$, $T=150$: The screening is neither weak nor strong. But the electrons are highly degenerate, the ^{16}O nuclei are rare and their charge at least fairly large compared with that of an α -particle. The conditions of Section VI are then satisfied. Equation (33) gives a value of 1.64 for the parameter F . Figure 2 then gives a value of 0.65 for $Y(0)/F$. Since $Z_2=z$ we find $-U_0/kT=Y(0)\approx 1.1$.

(iii) $\rho=10^8$, $T=150$: Again using (33) and Figure 2, we find $F=16.4$ and $-U_0/kT=6.1$.

VIII. REFERENCES

- BETHE, H. A. (1939).—*Phys. Rev.* **55**: 434.
 DEBYE, P., and HÜCKEL, E. (1923).—*Phys. Z.* **24**: 185.
 FOWLER, R., and GUGGENHEIM, E. A. (1949).—"Statistical Thermodynamics." Ch. 9. (Cambridge Univ. Press.)
 GAMOW, G., and CRITCHFIELD, C. L. (1949).—"Theory of Atomic Nucleus." p. 264. (Clarendon Press: Oxford.)
 KELLER, G. (1953).—*Astrophys. J.* **118**: 142.
 KELLER, G., and MEYEROTT, R. E. (1952).—Argonne Nat. Lab. 4771 and 4856.
 McDougall, J., and Stoner, E. C. (1938).—*Phil. Trans. A* **237**: 81.
 MAYER, J. E., and MAYER, M. G. (1940).—"Statistical Mechanics." Ch. 16. (J. Wiley & Sons: New York.)
 SALPETER, E. E. (1953).—*Annu. Rev. Nuclear Sci.* **2**: 41.
 SCHATZMAN, E. (1948).—*J. Phys. Radium* **9**: 46.
 SCHATZMAN, E. (1954).—*Astrophys. J.* **119**: 464.

OPTICAL DIFFRACTION EFFECTS PRODUCED BY AMPLITUDE AND PHASE CHANGES IN THE WAVE FRONT

By G. H. GODFREY*

[Manuscript received May 10, 1954]

Summary

The diffraction which occurs when certain areas of the wave front undergo changes of amplitude or phase or both, relative to the rest of the front, may be readily treated by a method which is virtually a generalization of Babinet's principle. Applications to both Fraunhofer and Fresnel diffraction are given and include phase gratings, striations, haloes, phase contrast techniques, zone plates, and holograms.

I. GENERAL PRINCIPLES

In the general Fraunhofer diffraction problem, a spherical light wave converging on or diverging from a point passes through an aperture. For the present purposes a simple aperture will be defined as one or more transparent areas in an opaque screen; and the real or virtual diffraction pattern on a plane through the centre of convergence or divergence is assumed known. The transparent and opaque parts of such an aperture will be referred to by their areas σ_1 and σ_2 , and the region intercepted by the full cone of rays by its area σ . If, in the general case now to be considered, the area σ is again divided into two parts σ_1 and σ_2 , either or both of which may transmit light with changes of amplitude and phase, the arrangement will be called a phase amplitude aperture. Let σ_1 and σ_2 transmit fractions r_1^2 and r_2^2 respectively of the light incident on them, and σ_1 produce a lag in phase φ relative to σ_2 . Particular cases are the pure phase aperture, $r_1=r_2=1$, $\varphi \neq 0$, and the pure amplitude aperture, $\varphi=0$. It will be shown that the diffraction pattern due to the phase amplitude aperture can now be expressed in terms of the corresponding simple aperture.

Suppose the wave functions due to the simple aperture, to its complement, and to the aperture of the lens be represented by

$$\Omega_1 = A_1 e^{-i\Delta_1}, \quad \Omega_2 = A_2 e^{-i\Delta_2}, \quad \Omega = A e^{-i\Delta}, \quad \dots \dots \dots (1)$$

where A_1 , A_2 , and A are amplitude factors and Δ_1 , Δ_2 , and Δ the corresponding phases. Then

$$\Omega = \Omega_1 + \Omega_2. \quad \dots \dots \dots (2)$$

With modifications of amplitude and phase the equations corresponding to (1) become

$$\Omega'_1 = r_1 A_1 e^{-i(\Delta_1 + \varphi)}, \quad \Omega'_2 = r_2 A_2 e^{-i\Delta_2}, \quad \Omega' = A' e^{-i\Delta'}, \quad \dots \dots (3)$$

where Ω' is the wave function for the phase aperture. Then

$$\Omega' = \Omega'_1 + \Omega'_2, \quad \dots \dots \dots (4)$$

* Division of Physics, C.S.I.R.O., University Grounds, Sydney.

and by elimination of $r_2 A_2$ and Δ_2 with the aid of (1), (2), (3), and (4)

$$\Omega' = A'e^{-i\Delta'} = r_1 A_1 e^{-i(\Delta_1 + \varphi)} + r_2 (Ae^{-i\Delta} - A_1 e^{-i\Delta_1}), \quad \dots (5)$$

or

$$A'e^{-i(\Delta' - \Delta_1)} = r_2 A e^{-i(\Delta - \Delta_1)} - A_1 (r_2 - r_1 e^{-i\varphi}). \quad \dots (6)$$

From (6)

$$\left. \begin{aligned} A' \cos (\Delta' - \Delta_1) &= r_2 A \cos (\Delta - \Delta_1) - r_2 A_1 + r_1 A_1 \cos \varphi, \\ A' \sin (\Delta' - \Delta_1) &= r_2 A \sin (\Delta - \Delta_1) + r_1 A_1 \sin \varphi, \end{aligned} \right\} \dots (7)$$

and

$$(A')^2 = r_2^2 A^2 + (r_1^2 + r_2^2 - 2r_1 r_2 \cos \varphi) A_1^2 - 2\{r_2^2 \cos (\Delta - \Delta_1) - r_1 r_2 \cos (\Delta - \Delta_1 - \varphi)\} A A_1 \quad \dots (8)$$

Let $C = A \cos \Delta$, $S = A \sin \Delta$, $C_1 = A_1 \cos \Delta_1$, $S_1 = A_1 \sin \Delta_1$. Then (8) becomes

$$\begin{aligned} (A')^2 &= r_2^2 A^2 + (r_1^2 + r_2^2 - 2r_1 r_2 \cos \varphi) A_1^2 - 2(r_2^2 - r_1 r_2 \cos \varphi)(CC_1 + SS_1) \\ &\quad + 2r_1 r_2 (C_1 S - S_1 C) \sin \varphi. \quad \dots (9) \end{aligned}$$

C , S , C_1 , and S_1 are proportional to the integrals employed in the theory of diffraction, and usually expressed by the same symbols, for calculating the illumination at a given point of a diffraction pattern. They are well known for a rectangle, a circle, a diffraction grating, and other simple cases. In terms of the illuminations E' and E_1 , (9) may be written

$$\begin{aligned} E'/E_1 &= A'^2/A_1^2 = (r_1^2 + r_2^2 - 2r_1 r_2 \cos \varphi) + \{r_2^2 A^2 - 2(r_2^2 - r_1 r_2 \cos \varphi)(CC_1 + SS_1) \\ &\quad + 2r_1 r_2 \sin \varphi(C_1 S - S_1 C)\}/A_1^2. \quad \dots (10) \end{aligned}$$

When $C = S = A = 0$, that is, at points where the diffraction pattern due to the lens alone has minimum (zero) values or at points sufficiently far from the geometric image of the source for this assumption to hold, (10) gives

$$\frac{E'}{E_1} = \left(\frac{A'}{A_1}\right)^2 = r_1^2 + r_2^2 - 2r_1 r_2 \cos \varphi. \quad \dots (11)$$

Equation (11) shows that, in the region to which it applies, the diffraction pattern due to a phase amplitude aperture is everywhere the same as that due to the corresponding simple aperture but with the illumination altered by the factor $(r_1^2 + r_2^2 - 2r_1 r_2 \cos \varphi)$. Thus for a pure phase aperture with $\varphi = \pi$ there is a fourfold amplification. Babinet's principle in its simple form is a particular case obtained by putting $r_1 = 0$, $r_2 = 1$.

The above results also apply to Fresnel diffraction patterns, C and S then being the corresponding integrals expressible in terms of Fresnel integrals. However, equation (11) is then restricted in application to the same extent as Babinet's principle when used in connexion with Fresnel diffraction (Wood 1911, p. 239).

It may be noted that, in the general case in which there are n phase amplitude apertures σ_s introducing phase changes φ_s ($s = 1, 2, \dots, n$) relative to a common background or reference aperture and absolute amplitude changes r_s , equation (5) is replaced by

$$\Omega' = A'e^{-i\Delta'} = r_c A e^{-i\Delta} + \sum_{s=1}^n A_s e^{-i\Delta_s} (r_s e^{-i\varphi_s} - r_c), \quad \dots (12)$$

where r_c is the amplitude change introduced by the background. This is equivalent to

$$A'e^{-i\Delta'} = r_c A e^{-i\Delta} + \sum_{s=1}^n q_s A_s e^{-i(\theta_s + \Delta_s)}, \quad \dots \quad (13)$$

where

$$\left. \begin{aligned} q_s \sin \theta_s &= r_s \sin \varphi_s, \\ q_s \cos \theta_s &= r_s \cos \varphi_s - r_c. \end{aligned} \right\} \quad \dots \quad (14)$$

Thus

$$\left. \begin{aligned} A' \cos \Delta' &= r_c A \cos \Delta + A_\sigma \cos \Delta_\sigma, \\ A' \sin \Delta' &= r_c A \sin \Delta + A_\sigma \sin \Delta_\sigma, \end{aligned} \right\} \quad \dots \quad (15)$$

giving

$$(A')^2 = r_c^2 A^2 + A_\sigma^2 + 2r_c A A_\sigma \cos(\Delta - \Delta_\sigma), \quad \dots \quad (16)$$

where

$$\left. \begin{aligned} A' \cos \Delta_\sigma &= \sum_{s=1}^n A_s q_s \cos(\theta_s + \Delta_s), \\ A' \sin \Delta_\sigma &= \sum_{s=1}^n A_s q_s \sin(\theta_s + \Delta_s), \end{aligned} \right\} \quad \dots \quad (17)$$

and

$$(A')^2 = \sum_{s=1}^n A_s^2 q_s^2 + \sum_{\substack{s,k=1 \\ s \neq k}}^n A_s A_k q_s q_k \cos(\theta_s - \theta_k + \Delta_s - \Delta_k). \quad \dots \quad (18)$$

Two particular cases of interest, which will be referred to later, may be mentioned. When all the apertures are pure phase apertures so that $r_s = r_c = 1$ equation (12) takes the form

$$A'e^{-i\Delta'} = A e^{-i\Delta} + 2 \sum_{s=1}^{s=n} A_s \left| \sin \frac{\varphi_s}{2} \right| e^{-i(\Delta_s + \varphi_s/2 \pm \pi/2)}, \quad \dots \quad (19)$$

the upper or lower sign being taken according as $\sin(\varphi_s/2)$ is positive or negative.

When all the apertures are pure amplitude apertures so that $\varphi_s = 0$, we have

$$A'e^{-i\Delta'} = r_c A e^{-i\Delta} + \sum_{s=1}^{s=n} A_s |r_s - r_c| e^{-i(\Delta_s + \alpha)}, \quad \dots \quad (20)$$

where $\alpha = 0$ or π according as $r_s - r_c$ is positive or negative. If for one of the phase amplitude apertures σ_p the phase and amplitude factors φ_p and r_p vary continuously over its area it is obvious that the contribution of this aperture to the summation term in (12) is

$$\int P e^{-i\Delta_p} (r_p e^{-i\varphi_p} - r_c) d\sigma_p,$$

where the integration is taken over the area σ_p and P and Δ_p are functions of the position of the element $d\sigma_p$ and of the point at which Ω' is required.

II. APPLICATIONS

In applying equations (9) or (10) to any particular problem it is only necessary to know the values of the integrals C , S , C_1 , and S_1 (which correspond to the case in which r_1 and r_2 are unity and φ is zero). However, it should be

pointed out in connexion with Fraunhofer diffraction that the integrals C_1 and S_1 depend on the position of the area σ_1 as well as its shape and orientation but in such a way that a displacement of σ_1 in its own plane without rotation does not alter $C_1^2 + S_1^2$. Thus, as is well known, such a displacement does not alter the diffraction pattern of a simple aperture except as regards the phase Δ_1 associated with it. In the case of a phase amplitude aperture, however, the pattern is modified by such a displacement (except at points at which equation (11) applies), as is seen from (9). It should be noted further that when an aperture is symmetrically placed with regard to the incident wave the integral S or S_1 , as the case may be, is everywhere zero since it represents the integral of an odd function. At the geometric image, even when symmetry is absent, S and S_1 are zero and C and C_1 (or A and A_1) proportional respectively to the areas σ and σ_1 since all rays reach the focus in the same phase.

Examples will now be considered which have been worked out by other methods, generally from first principles, but which may be considered as simple applications of the above formulae.

(a) *The Phase Grating*

A phase or laminary grating (Wood 1911, p. 211) consists of a series of transparent strips, alternate strips being of equal width and producing a phase change differing from that of the other strips by a constant. When used to form spectra one set of alternate strips of such a grating represents collectively σ_1 and the other set the complementary area σ_2 . Application of (10) and (11) gives all the effects described qualitatively by Wood (1911, p. 211). In particular when $r_1 = r_2 = 1$ and for a particular wavelength for which $\varphi = \pi$ there is a fourfold intensification of the spectra at points where (11) applies. If, in addition, the widths of the strips are all equal the illumination of the central image is zero for this wavelength. These results are deduced, in the first place, for a point source and the principal axis of the pattern but they apply equally to the case of a line source parallel to the laminae.

(b) *Effects of Striae in a Lens on the Illumination in the Image Plane*

Striations in optical glass introduce a phase change in the light passing through them. If this phase change is uniform they would act as phase apertures in the sense used here. This assumption of uniformity may seem to introduce an over-simplification but we shall show that it is the mean value of φ over the striation which is important. Suppose a lens has a circular aperture of radius R and a single uniform striation in the form of a narrow transparent strip of length a and breadth b , so that $r_1 = r_2 = 1$, then $S = 0$ and $C = A$ at all points of the image plane. For a centrally placed striation,

$$\left. \begin{aligned} S_1 &= 0, \\ C_1 &= ab \left[\frac{\sin \frac{kb}{2}}{\frac{kb}{2}} \right] \end{aligned} \right\} \dots\dots\dots (21)$$

at points on the major axis of the diffraction pattern, k being given by

$$k = \frac{2\pi\theta}{\lambda},$$

where θ is the angle between the optic axis and the line joining the centre of the lens to the point in the image plane at which the illumination is required.

The values of C^2 at the geometric image, first minimum, and second maximum of the diffraction pattern due to the lens alone are proportional to 1, 0, and 0.017 and occur respectively at points for which θ has the values 0, $0.61 \lambda/R$, and $0.81 \lambda/R$. The corresponding values of S are zero. For a large striation such that $ab/R^2 = 0.1$ and $\varphi = \pi$ equations (9) and (21) give for the illumination at these three points the comparative figures 0.87, 0.004, 0.005. For $\varphi = 1$ radian and $ab/R^2 = 0.04$ the corresponding figures are 0.9873, 0.0002, and 0.0160. Françon (1948a), using a direct method applicable to his particular problem, considers the effect of a striation which introduces a phase change which varies with the displacement x from its major axis in accordance with the relation $\varphi \propto (1 + \cos 2\pi x/b)$. His corrected results (Françon 1948b) for a striation of the same dimensions as in the last example and producing the same average phase change are, for the three points mentioned, 0.9852, 0.0001, 0.0195, so that the effects are not very sensitive to the distribution of phase change across the striation.

The effect of a striation, however, depends to some extent upon its position, the diffraction pattern being unsymmetrical for a striation which is off centre except when $\varphi = \pi$. Thus in the first example, if the striation is displaced through a distance $R/2$ from the centre, the illumination at the first and second points is unaffected but at the third it is increased to 0.033.

(c) *Haloes etc.*

Consider a large number, n , of diffraction apertures, represented collectively by σ_1 , similar as regards shape and orientation, but randomly spaced. At points sufficiently far from the geometric image equation (11) is replaced by

$$E' = n(r_1^2 + r_2^2 - 2r_1r_2 \cos \varphi)e_1, \quad \dots \dots \dots (22)$$

where e_1 is the illumination produced by a single aperture. At the geometric image itself the illumination is given by

$$E' = r_2^2 E + (r_1^2 + r_2^2 - 2r_1r_2 \cos \varphi)ne_1 - 2(ne_1 E)^{1/2}(r_2^2 - r_1r_2 \cos \varphi). \quad \dots (23)$$

In the special case where $\varphi = \pi$ and each of the constituent areas is $1/2n$ of the area of the lens, the diffraction pattern is intensified at points away from the geometric image by a factor $4n$, as compared to the effect of a simple aperture, and the illumination at the geometric image is reduced to zero. The haloes obtained with transparent disks by R. W. Wood (1911, p. 254) may be treated in this way. A phase grating with random spacing behaves in a similar manner.

(d) *Phase Contrast Techniques* (Zernike 1942)

The object of these techniques is to render visible, by means of a microscope or similar optical arrangement, details which may be described as pure phase

apertures ($r_1=r_2=1$, $\varphi \neq 0$). Suppose that, instead of receiving the diffraction pattern of such an aperture (i.e. the detail to be examined) on the plane conjugate to the original point source, the wave characteristics (amplitude and phase) associated with the diffraction pattern are so modified in passing through this plane as to correspond at all points to those due to a pure amplitude aperture ($r_1 \neq r_2$, $\varphi=0$). Then an optical system, placed so as to receive the light passing through this plane and focused on the detail, will image it as such an amplitude aperture. The resulting image will have what is called bright contrast if the ratio $R=r_1/r_2$ is greater than unity and dark contrast if R is less than unity. If φ is small, the wave characteristics for the pure phase aperture in the off-centre part of the diffraction field where $A=0$ are, from (6), given by

$$A'e^{-i(\Delta'-\Delta)} = -i\varphi A_1 = |\varphi| A_1 e^{\mp i\pi/2}$$

in which the upper or lower signs are to be taken according as φ is positive or negative. Thus

$$A' = |\varphi| A_1, \quad \Delta' = \Delta_1 \pm \pi/2.$$

For the pure amplitude aperture in the same region we have, from (6),

$$A'e^{-i(\Delta'-\Delta)} = (r_1 - r_2) A_1$$

or

$$A' = |r_1 - r_2| A_1 \quad \text{and} \quad \Delta' = \Delta_1 \quad \text{or} \quad \Delta_1 + \pi,$$

according as $r_1 >$ or $< r_2$. Thus for a positive value of $r_1 - r_2$ the phase at all points of the off-centre diffraction pattern for the pure phase aperture is in advance of that for the pure amplitude aperture by $\pm \pi/2$ according as φ is positive or negative and (taking $3\pi/2$ as equivalent to $-\pi/2$) vice versa for $r_1 - r_2$ negative. Again the amplitude at all such points due to the pure phase aperture bears the constant ratio $|\varphi|/|r_1 - r_2|$ to that due to the pure amplitude aperture. Thus the required correspondence exists for any given value of φ , r_1 , and r_2 at all points in this part of the pattern.

For the central part close to the geometric image the phase due to both types of aperture is approximately Δ_1 or Δ and the amplitudes for the two cases respectively are, from first principles or by (6), approximately A and $r_1 A_1 + r_2 (A - A_1)$. Thus, to establish for all points of the pattern in the two cases a constant difference of phase and a constant ratio of amplitudes, it is necessary to introduce the small correcting plate used in phase contrast techniques to intercept the rays reaching the vicinity of the geometric image of the point source. This plate should produce a phase change in the transmitted light of $\pm \pi/2$ and have an amplitude transmittance t where

$$\frac{At}{r_1 A_1 + r_2 (A - A_1)} = \frac{|\varphi|}{|r_1 - r_2|}, \quad \dots \dots \dots (24)$$

or

$$\frac{\sigma t}{\sigma_1 R + (\sigma - \sigma_1)} = \frac{|\varphi|}{|R - 1|}, \quad \dots \dots \dots (25)$$

the contrast being bright or dark according as the phase change introduced is of the same or opposite sign to that of φ . The value of the ratio R determining the contrast is thus related to t by

$$R = \frac{t\sigma \pm \varphi(\sigma - \sigma_1)}{t\sigma \mp \varphi\sigma_1} \simeq 1 \pm \frac{\varphi}{t}, \quad \dots\dots\dots (26)$$

the upper or lower sign being taken according as the plate introduces a phase change of $+\pi/2$ or $-\pi/2$.

Maximum contrast is obtained when $R = \infty$ or 0 , i.e. when

$$t = \frac{\sigma_1}{\sigma} \varphi \quad \text{or} \quad t = \frac{\sigma - \sigma_1}{\sigma} \varphi, \quad \dots\dots\dots (27)$$

for bright and dark contrast respectively. For values of t large compared with these the approximate form of (26) holds.

When the detail to be examined consists of a number of pure phase apertures introducing small phase changes φ_s relative to a common background it is readily shown by means of equations (19) and (20) that (26) is replaced by

$$R_s = 1 \pm \frac{\varphi_s \sigma}{t\sigma \mp \sum \varphi_s \sigma_s} = 1 \pm \frac{\varphi_s \sigma}{t\sigma \mp \int \varphi d\sigma} \simeq 1 \pm \frac{\varphi_s}{t}. \quad \dots\dots (28)$$

The integral form for $\sum \varphi_s \sigma_s$ is of use when φ changes continuously in any part of the detail.

(e) *Zone Plate with Phase Reversal* (Rayleigh 1902)

The applications so far considered have been concerned with Fraunhofer diffraction. This and the next are examples involving Fresnel diffraction. For all such applications (except as mentioned in connexion with equation (11)) the aperture σ_2 extends effectively from the boundary of σ_1 to infinity. Similarly σ is a region of space of infinite extent.

For a zone plate illuminated by a divergent beam from a point source the light is so concentrated at the focal points of the plate that A is negligible compared to A_1 . Thus, if the opaque parts of the plate are rendered transparent but produce a relative phase change of π , substitution of $r_1 = r_2 = 1$, $\varphi = \pi$, $A = 0$ in (8) gives $(A')^2 = 4A_1^2$ so that the focal properties of the plate are intensified fourfold.

(f) *The Hologram* (Gabor 1949, 1951)

If a small object (σ_1, r_1, φ) be placed in a divergent wave from a point source the wave function associated with the diffraction pattern is obtained from (5) by setting $r_2 = 1$, namely,

$$\Omega' = A'e^{-i\Delta'} = Ae^{-i\Delta} - A_1\{e^{-i\Delta_1} - r_1 e^{-i(\Delta_1 + \varphi)}\}. \quad \dots\dots (29)$$

If σ_1 is sufficiently small, A_1^2 is negligible, so from (8)

$$A'^2 = A^2 - 2AA_1[\cos(\Delta - \Delta_1) - r_1 \cos\{\varphi - (\Delta - \Delta_1)\}]. \quad \dots (30)$$

Now suppose a photographic plate is exposed to the diffracted wave and then developed by reversal, or its equivalent, with an overall gamma of 2. Then the

developed plate or hologram, as it is called, will have an amplitude transmission proportional to $(A')^2$. If the plate is replaced in its original position but with the object removed, so that the direct light from the source falls on it, the wave function Ω'' in the plane immediately behind the emulsion will, except for a constant of proportionality, be represented by

$$\Omega'' = A'^2 A e^{-i\Delta} = A^2 [A e^{-i\Delta} - A_1 \{e^{-i\Delta_1} - r_1 e^{-i(\varphi + \Delta_1)} + e^{-i(2\Delta - \Delta_1)} - r_1 e^{i(\varphi - 2\Delta + \Delta_1)}\}], \quad \dots\dots\dots (31)$$

that is, by

$$\Omega'' = A^2 \Omega' - A^2 A_1 \{e^{-i(2\Delta - \Delta_1)} - r_1 e^{i(\varphi - 2\Delta + \Delta_1)}\}, \quad \dots\dots\dots (32)$$

or

$$\Omega'' = A^2 e^{-2i\Delta} \Omega'^* - A^2 A_1 \{e^{-i\Delta_1} - r_1 e^{-i(\varphi + \Delta_1)}\}, \quad \dots\dots\dots (33)$$

where Ω'^* is the conjugate of Ω' .

The interpretation of (32) is simple. If it were possible to disregard all but the first term on the right, Ω'' would be proportional to Ω' for a uniform source. Thus from Huyghens' principle the disturbance everywhere behind the plate would be just as though the object were restored to position and the plate removed; in other words, in the original location of the object there would be a virtual image which could be viewed with a suitable optical system. Actually what is seen is an imperfect image the defects being due to the remaining terms in (32).

To interpret the alternative expression given by (33) we again consider only the first term on the right of the equation. It should then be noted that in both (32) and (33) the time or phase origin is the same. Treating first the case in which the object is initially nearer to the source than to the photographic plate, suppose the source to be formed by a wave converging on the point where it is located from the side remote from the plate. Consider now an object in the converging beam at a position which is the mirror image of the original object (σ_1, r_1, φ) in reference to a spherical surface located in the plane of the photographic plate and concentric with the source. Then it may be shown (Appendix I) that this object has the same diffraction pattern (as regards amplitude) in the plane of the plate, and hence the same hologram, as the original object. Thus the hologram reconstruction must contain as an alternative to the virtual image described by (32) a second inverted virtual image behind the source in this position. When phase relations are taken into account the complete expression for the wave function associated with the diffraction pattern of the second object (Appendix I, equation (A4)) is $A^2 e^{-2i\Delta} \Omega'^*$, thus accounting for alternative (33). When the original object is closer to the plate than to the source a similar argument applies. However, in this case the mirror image which constitutes the secondary object is on the side of the plate remote from the source and this object would produce the same diffraction pattern in the plane of the plate with a light beam converging on the position of the original point source from the opposite direction. The wave function associated with this pattern (Appendix I, equation (A7)) is $A^2 e^{2i\Delta} \Omega'$. The conjugate quantity $A^2 e^{-2i\Delta} \Omega'^*$ occurring in (33) now represents a wave front leaving the hologram

and converging on the location of the secondary object to form a real image of the original object. This image may be viewed by means of a short focus optical system such as a microscope with the aid, if necessary, of phase contrast (or similar techniques). Again the neglected terms in (33) represent defects of the image which are thus seen to be present in the case of all the images reproduced by the hologram.

(g) Other Applications

Zernike (1948) has described the Fraunhofer diffraction pattern due to a slit covered with a transparent layer of strongly absorbing metal with a narrow scratch in the centre of the slit. Obviously equation (8) contains the complete theory of this experiment, the amplitude factors A and A_1 referring respectively to the slit and scratch as simple apertures. The effect of varying the slit width and the position of the scratch may easily be observed with the aid of an optical bench.

Zernike (1948) has also described the existence of fringes within the geometrical shadow of a straight edge when the diffracting screen is slightly transparent. Again referring to equation (8), A and Δ relate in this case to the unimpeded wave from the source and A_1 and Δ_1 to the pattern produced by the corresponding opaque screen. Thus within the shadow the term in A_1^2 is negligible and the fringes referred to arise from the variations in the coefficient of AA_1 . For example, if the effect of φ is neglected this coefficient is proportional to $\cos(\Delta - \Delta_1)$ or $\sin \Delta_1$ since Δ has the value $\pi/2$ (Preston 1928). The variations of $\sin \Delta_1$ are readily obtainable from Cornu's spiral or a table of Fresnel's integrals.

Similar but modified effects to those just mentioned are produced by diffraction at the edge of a transparent lamina when phase changes are introduced (Wood 1911, p. 250) and the same equation gives the analysis.

III. REFERENCES

- FRANÇON, M. (1948a).—*Rev. Opt. (Théor. Instrum.)* **27**: 595.
 FRANÇON, M. (1948b).—*Rev. Opt. (Théor. Instrum.)* **27**: 761.
 GABOR, D. (1949).—*Proc. Roy. Soc. A* **197**: 454.
 GABOR, D. (1951).—*Proc. Phys. Soc. Lond.* **B 64**: 449.
 PRESTON, T. (1928).—"Theory of Light." 5th Ed. p. 300. (Macmillan and Co.: London.)
 RAYLEIGH, LORD (1902).—"Scientific Papers." 1st Ed. Vol. 3, p. 78. (Cambridge Univ. Press.)
 WOOD, R. W. (1911).—"Physical Optics." 2nd Ed. (The Macmillan Co.: New York.)
 ZERNIKE, F. (1942).—*Physica* **9**: 686, 974.
 ZERNIKE, F. (1948).—*Proc. Phys. Soc. Lond.* **61**: 158.

APPENDIX I

Equivalent Diffraction Apertures

Suppose light converges on the point O (Fig. 1) in the direction $B'O$ and is diffracted by an aperture AB . It will be shown that, if MX is a spherical surface concentric with O and further from AB than is O , the diffraction pattern produced on MX is the same as would be produced by an aperture $A'B'$ which is the mirror image of AB in respect of the surface MX .

Consider two conjugate points P and P' . Since all optical paths from P to P' must be equivalent we have for any point X

$$PX + P'X = PM + P'M,$$

or

$$P'M - P'X = PX - PM. \quad \dots\dots\dots (\text{A1})$$

Thus light from P' reaches X in a phase which is as much ahead of the phase in which it reaches M as the phase with which light from P reaching X is behind the phase with which it reaches M . Let the absolute value of this difference be Δ_n . Further, let the phase of light on reaching M along the path $POP'M$ lag behind that at O by an amount Δ . Then the phase with which the light diffracted by an infinitesimal area at P will reach X is $\Delta - \Delta_n$, whilst the phase

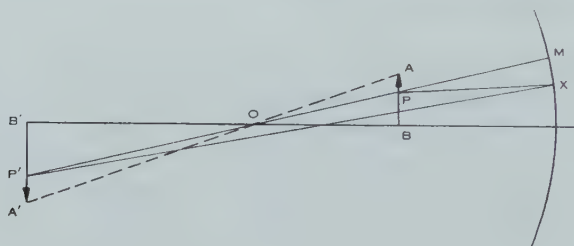


Fig. 1

with which the light diffracted by the conjugate element at P' will reach X is $\Delta + \Delta_n$, Δ being a constant for all positions of P and P' . The amplitudes of the diffracted light reaching X from the two conjugate elements at P and P' respectively are equal since the flux through them is the same. Denoting these amplitudes by a_n , the resultant wave functions at X due to the apertures AB and $A'B'$ are represented by

$$\Omega = \sum a_n e^{-i(\Delta - \Delta_n)} = e^{-i\Delta} \sum a_n e^{i\Delta_n}, \quad \dots\dots\dots (\text{A2})$$

$$\Omega' = \sum a_n e^{-i(\Delta + \Delta_n)} = e^{-i\Delta} \sum a_n e^{-i\Delta_n}. \quad \dots\dots\dots (\text{A3})$$

Therefore

$$\Omega' = e^{-2i\Delta} \Omega^*, \quad \dots\dots\dots (\text{A4})$$

where Ω^* is the conjugate of Ω .

Thus the amplitudes of Ω' and Ω are the same at all points and the diffraction patterns due to the two apertures are identical in shape and position. Obviously these relations are reciprocal, that is, equation (A4) holds when AB and its image are interchanged.

When the aperture AB is closer to MX than to O as in Figure 2, a similar conclusion may be drawn but with the following modification. The aperture AB gives the same diffraction pattern on MX , for light diverging from O , as its image $A'B'$ for light converging on O . To prove this we note that equation (A1) is replaced by

$$P'X - P'M = PX - PM,$$

which follows from the fact that waves from P reach M and X respectively with the same phase difference as if they had originated from the image P' . Thus, if we consider two wave trains, one converging on O and the other diverging from it, then, at the instant when both trains are in phase at O , the phase of both at M will be the same (Δ say); but at the same instant the phase at X of the waves diffracted by P and P' will be $(\Delta - \Delta_n)$ and $(\Delta + \Delta_n)$ where Δ_n corresponds to the common path difference.

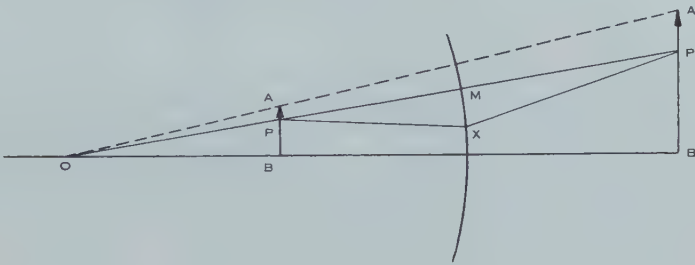


Fig. 2

Hence in place of (A2) and (A3) we have for the diverging wave

$$\Omega = \sum a_n e^{-i(\Delta - \Delta_n)} = e^{-i\Delta} \sum a_n e^{i\Delta_n}, \quad \text{..... (A5)}$$

and for the converging wave

$$\Omega' = \sum a_n e^{i(\Delta + \Delta_n)} = e^{i\Delta} \sum a_n e^{i\Delta_n}, \quad \text{..... (A6)}$$

or

$$\Omega' = e^{2i\Delta} \Omega. \quad \text{..... (A7)}$$

It is obvious from (A7) that $A'B'$ and AB produce the same diffraction pattern on MX .

It may be pointed out that the conjugate of Ω' as defined by (A7) would represent a wave front converging on $A'B'$ to form a real image of AB . If a device could be placed in position MX which would impart to a beam from O transmitted through it a wave front defined by $(\Omega')^*$ such a real image would be formed. The hologram succeeds partially in producing this effect.

When MX is sufficiently distant from O the diffraction screen MX may be plane. On the other hand when the diffraction screen passes through O (Fraunhofer diffraction) the above principles cease to apply.

THE INFLUENCE OF THE CATHODE MATERIAL ON MEASURED BREAKDOWN STRENGTHS OF SOLID AND LIQUID DIELECTRICS*

By J. J. O'DWYER†

[Manuscript received April 5, 1954]

Summary

An outline is given of a general method of accounting for the effect of electronic and ionic space charge in the breakdown of solid and liquid dielectrics. Detailed calculations are performed for a special model of both a liquid and a solid dielectric, and the results are compared with recent experimental work. Exact calculation is not possible, but it is shown that for some substances an effect due to cathode material may well be measurable, and the factors influencing such an effect are examined.

I. INTRODUCTION

The intrinsic breakdown strength of certain dielectric materials has been calculated by Fröhlich (1937, 1939, 1947). The materials treated have been ionic crystals in the "low temperature" region, and "amorphous dielectrics". The latter term refers to the model assumed for the electronic structure of the dielectric and is taken to include ionic crystals in the "high temperature" region and various solids and liquids of more complicated structure. Attempts to measure this intrinsic breakdown strength experimentally are complicated by the fact that the theory takes no account of the finite size of the dielectric specimen, the finite rise time of the applied voltage, or the effect of electrode materials.

Various authors have discussed this latter effect, but opinions differ both as to its existence and to its mechanism presuming that it exists. Oakes (1948) obtained a significant difference between measurements on the electric strength of polythene using silver and graphite cathodes. von Hippel and Alger (1949) have performed experiments on the breakdown strength of potassium bromide (in its high temperature region) with gold, mercury, steel, and potassium bromide solution as cathodes, and claim much larger effects of this kind; it should be mentioned, however, that in the experiments by von Hippel and Alger the specimens with different electrode material were not conditioned in the same way so that the interpretation of the differences as due to electrode material cannot be considered as established. More recently Calderwood, Cooper, and Wallace (1953) have measured the breakdown strength of ionic crystals which were very carefully prepared and annealed. They found no significant difference in the breakdown strength of potassium chloride at room temperature using alternately graphite and silver electrodes; and a difference that was just about significant using graphite and potassium chloride solution alternately as cathodes.

* Based in part on E.R.A. Report L/T 263.

† Division of Electrotechnology, C.S.I.R.O., University Grounds, Sydney.

Similarly, Cooper and Grossart (1953) found no significant difference in the breakdown strength of potassium bromide using gold, graphite, and potassium bromide solution alternately as cathodes. This is in opposition to the earlier reported results of von Hippel and Alger.

Similarly, experimental results on the breakdown of dielectric liquids are conflicting. Various workers have measured the effect of different cathodes on the breakdown strength of *n*-hexane. Using a conventional sphere-sphere electrode system Edwards (1951), Salvage (1951), and Goodwin and Macfadyen (1953) all report the existence of such an effect while Lewis (1953*a*) reports that the effect does not exist. However, in further experiments using a point-plane electrode system, Lewis (1953*b*) found that the cathode material influenced the breakdown strength if the plane was the cathode, but that there was no measurable effect if the point was the cathode. The cathodes used in these experiments have been aluminium, chromium, silver, nickel, platinum, phosphor-bronze, and steel, and in all cases in which the existence of the effect is reported the measured breakdown strength increased with increasing cathode work function.

It is the purpose of the present work to interpret the possible influence of the cathode material on measured breakdown strengths, on the assumption that each material possesses an intrinsic electric strength. In the calculations below it is supposed that the specimen of dielectric under test is placed between two parallel metal electrodes and the applied potential altered so slowly that the charge distribution may always be assumed to be in a steady state. In approaching this steady state a space charge will have been built up in the dielectric, which will be so distributed as to give continuity of current across the specimen. The variation of potential will not be linear, and under certain conditions the maximum value of the field strength may be considerably greater than the average field strength. It is assumed that breakdown will occur when this maximum field strength exceeds the intrinsic breakdown strength.

II. GENERAL METHOD OF SOLUTION FOR CERTAIN SOLID DIELECTRICS

We consider now solid dielectrics in which the current is purely electronic. Continuity of current requires that the current of electrons leaving at the anode be equal to that entering at the cathode; the former is determined by the conductivity of the dielectric near the anode, and the latter by cold emission from the cathode into the conduction band of the dielectric. This can be expressed by

$$I_c(F_c, \alpha) = \sigma_a F_a, \quad \dots \quad (1)$$

where F_c and F_a are the fields near the cathode and anode respectively, α represents parameters of the metal and the dielectric which are relevant to cold emission, and σ_a is the conductivity of the dielectric near the anode. The electronic space charge will be distributed to give the highest field strength near the anode and the lowest near the cathode. Near breakdown (1) thus becomes

$$I_c(F_c^*, \alpha) = \sigma^* F^*,$$

so that

$$F_c^* = F_c^*(\alpha, \sigma^*, F^*), \quad \dots \quad (2)$$

thus determining the field near the cathode just before breakdown. The quantities σ^* and F^* are the conductivity and field strength just before intrinsic breakdown.

The conductivity of the dielectric will be increased by an electronic space charge, or a strong field applied to it, or both (cf. Fröhlich 1947) and we may write

$$\sigma = \sigma(\sigma_0, n, F, \beta), \quad \dots \dots \dots (3)$$

where σ_0 is the conductivity in the absence of a field, n is the space charge electron density, and β represents parameters of the dielectric.

Continuity of current gives

$$\begin{aligned} F\sigma &= \text{const.} \\ &= F^*\sigma^*, \quad \dots \dots \dots (4) \end{aligned}$$

so from equations (3) and (4) we have

$$n = n(\sigma_0, \beta, F, F^*). \quad \dots \dots \dots (5)$$

Poisson's equation can be written

$$\frac{dF}{dx} = 4\pi ne, \quad \dots \dots \dots (6)$$

where x is the position coordinate measured from the cathode. The variables are separable, so that integration of (6) gives

$$F = F(x, \beta, F^*, F_c^*) \quad \dots \dots \dots (7)$$

with the use of the boundary condition (equation (2)). If the anode and cathode are a distance d apart, then we can write for the apparent fractional decrease of breakdown strength

$$\left. \begin{aligned} D &= 1 - \frac{F_0}{F^*}, \\ F_0 &= \frac{\int_0^d F dx}{d}. \end{aligned} \right\} \quad \dots \dots \dots (8)$$

where

III. CALCULATIONS FOR AN AMORPHOUS DIELECTRIC

Since the detailed manner of variation of the number of conduction electrons with field strength is not known for a dielectric in its low temperature region (due to unsolved problems connected with internal ionization and recombination processes), we shall perform the explicit calculations for the case of an amorphous dielectric only. In this sense the term "amorphous" is taken to refer to the model of the electronic structure introduced by Fröhlich (1947). The electronic energy levels consist of a continuum of conduction levels immediately below which isolated shallow traps cover a range ΔV ; there are in addition deep traps whose energy is an amount W below the lowest of the shallow traps. For this model, the conductivity as a function of field strength has been calculated by Fröhlich, and his results will be required below.

Using a method similar to that given by Mott and Sneddon (1949) we can derive a Fowler-Nordheim type equation for the current emitted by the cathode into the conduction levels of the dielectric as

$$I_c = \frac{e^3 F_c^2}{8\pi h \varphi} \exp \left(-\frac{4}{3} K \varphi^{3/2} / e F_c \right), \quad \dots \dots \dots (9)$$

where $K^2 = 8m\pi^2/h^2$, e and m are the electronic charge and mass respectively, h is Planck's constant, and φ is the amount by which the energy of the lowest conduction level in the dielectric exceeds the energy of the top of the Fermi distribution in the metal. From equation (2) then

$$\frac{e^3 F_c^{*2}}{8\pi h \varphi} \exp \left(-\frac{4}{3} K \varphi^{3/2} / e F_c^* \right) = \sigma^* F^*, \quad \dots \dots \dots (2a)$$

from which we can now find F_c^* , which is required as a boundary condition for the solution of the differential equation.

The number of electrons in conduction levels in the dielectric can be written

$$N_c(T) = k_1(T)(N+n) \exp \left(-\frac{W}{kT} \right), \quad \dots \dots \dots (10)$$

where T is the electronic temperature (assuming that electron-electron collisions are so much more frequent than electron-lattice collisions that an electronic temperature is attained—not necessarily the same as the lattice temperature T_0), k_1 is a factor which varies slowly with T compared with the exponential term, N is the number of trapped and conduction electrons, and n the number of space charge electrons. (It is assumed in the derivation of (10) that N is very much less than the total number of deep traps (cf. Mott and Gurney 1949).)

If there is no field and no space charge the electronic temperature will equal the lattice temperature and

$$N_c(T_0) = k_1(T_0)N \exp \left(-\frac{W}{kT_0} \right). \quad \dots \dots \dots (11)$$

Combining equations (10) and (11) and treating k_1 as constant, we have for the ratio of the conductivities

$$\frac{\sigma}{\sigma_0} = \left(1 + \frac{n}{N} \right) \exp \left(-\frac{W}{kT} + \frac{W}{kT_0} \right). \quad \dots \dots \dots (12)$$

Fröhlich (1947) has shown that

$$\frac{W}{kT_0} - \frac{W}{kT_1} \simeq \frac{F^2}{F^{*2}} \frac{W}{\Delta V}, \quad \dots \dots \dots (13)$$

provided that $F \lesssim F^*$ and where T_1 is the steady electronic temperature built up under the influence of a field F . Equations (12) and (13) then give

$$\frac{\sigma}{\sigma_0} = \left(1 + \frac{n}{N} \right) \exp \left(\frac{W}{\Delta V} \frac{F^2}{F^{*2}} \right). \quad \dots \dots \dots (3a)$$

Substituting (3a) into (4) gives at once

$$\frac{n}{N} = \frac{F^*}{F} \exp \left\{ \frac{W}{\Delta V} \left(1 - \frac{F^2}{F^{*2}} \right) \right\} - 1. \quad \dots \dots \dots (5a)$$

Using (5a) in (6) and separating the variables we find

$$x = \frac{1}{4\pi Ne} \int \frac{dF}{\frac{F^*}{F} \exp \left\{ \frac{W}{\Delta V} \left(1 - \frac{F^2}{F^{*2}} \right) \right\} - 1}.$$

For values of $F \lesssim F^*$ the exponential can be well approximated by two terms and elementary integration gives

$$x = \frac{\Delta V}{W} \frac{F^*}{4\pi Ne} \left[-\ln \left(\frac{F}{F^*} - 1 \right) + \left(\frac{\Delta V}{W} + 1 \right) \ln \left(\frac{F}{F^*} + \frac{\Delta V}{W} + 1 \right) \right]. \quad (14)$$

The quantity $(1 + \Delta V/W)$ is of order unity, so for $F \lesssim F^*$ the first term of (14) will be very much greater than the second. Neglecting this second term and applying the boundary condition $F = F_c^*$ at $x = 0$ when breakdown is about to commence at the anode, we have

$$\frac{F}{F^*} = 1 - \left(\frac{F^* - F_c^*}{F^*} \right) \exp \left(- \frac{4\pi Ne}{F^*} \frac{W}{\Delta V} x \right), \quad (7a)$$

which gives the field strength at any point when the dielectric is about to break down. Performing the averaging and substituting in (8), we have for the apparent fractional decrease of breakdown strength

$$\begin{aligned} D &= \frac{F^* - F_c^*}{4\pi Ned} \frac{\Delta V}{W} \left[1 - \exp \left(- \frac{4\pi Ned}{F^*} \frac{W}{\Delta V} \right) \right] \\ &= \frac{F^* - F_c^*}{\beta} \left[1 - \exp \left(- \frac{\beta}{F^*} \right) \right], \quad (8a) \end{aligned}$$

where we have written

$$\beta = 4\pi Ned \frac{W}{\Delta V} \quad (15)$$

as a parameter which has the dimensions of a field strength and is characteristic of the specimen under test.

IV. GENERAL METHOD OF SOLUTION FOR LIQUID DIELECTRICS

We consider now a liquid dielectric in which the electronic current emitted from the cathode is

$$I_c = I_c(F_c, \alpha), \quad (16)$$

where again α represents parameters of the metal and dielectric which are relevant to cold emission.

The liquid molecules will be ionized by the electrons and the steady state field distortion will be due to the space charge of both electrons and positive ions. Poisson's equation will then be found to take the form

$$\frac{dF}{dx} = \frac{dF}{dx}(I_c, \beta, F, x), \quad (17)$$

where β again represents parameters of the dielectric specimen under test (e.g. its size and shape, the mobilities of electrons and positive ions, and ionization and recombination coefficients). If (17) can be integrated the result will be

$$F = F(I_c, \beta, F_c, x), \quad \dots \quad (18)$$

where $F = F_c$ when $x = 0$ has been introduced as a boundary condition. Using (16) in (18), we have immediately

$$F = F(\alpha, \beta, F_c, x). \quad \dots \quad (19)$$

In the case of the liquid dielectric the lower mobility of the positive ions as compared with the electrons will result in the field strength being greatest near the cathode and falling away towards the anode. Breakdown will thus occur when $F_c = F^*$ so that the breakdown field at any point is

$$F = F(\alpha, \beta, F^*, x). \quad \dots \quad (20)$$

The apparent fractional decrease in breakdown strength is then given by equation (8) as previously.*

V. CALCULATIONS FOR A LIQUID DIELECTRIC

The case of a liquid dielectric can also be made amenable to calculation if certain assumptions are made. It will be assumed that the electron current I_c from the cathode produces positive ions by collision-ionization with a multiplication coefficient η which will be strongly dependent on the field strength. Recombination will be neglected as being negligible compared with ionization. If in addition the mobility of the electrons is very much greater than the mobility of the positive ions it can be shown that Poisson's equation gives (cf. Loeb 1939)

$$F \frac{dF}{dx} = - \frac{4\pi I_c}{k_+} \left\{ \exp \left(\int_0^d \eta dx \right) - \exp \left(\int_0^x \eta dx \right) \right\},$$

where k_+ is the mobility of the positive ions. (This equation is standard for the case of discharge in a gas.) It can be taken that

$$\eta = \eta_0 F^2 \quad \dots \quad (21)$$

* Goodwin and Macfadyen (1953) combine equations (16) and (18) to produce a breakdown criterion. They do this by showing that, for sufficiently large F_0 (average field strength across the specimen), no simultaneous solution for I_c and F_c can be obtained from these equations. This procedure, however, seems unsound for several reasons. Firstly, the possibility that the liquid possesses an intrinsic breakdown strength is ignored, and in view of the relative constancy of measured breakdown strengths it seems very probable that the concept of the intrinsic breakdown strength applies to a liquid. This would be calculated as that field strength for which no equilibrium is possible between the rate at which the charge carriers in the liquid receive energy from the field and the rate at which they can transfer this excess energy to the bulk of the liquid. Secondly, the constants used by these authors in the Fowler-Nordheim equation correspond to values of ϕ for which not cold emission but thermionic emission would be the dominant effect. This is in the main due to the method by which they determine these constants from experiment by extrapolation of current v . gap length curves (for a given average field strength) to zero gap length. This is not permissible since the field distortion near the cathode is such that the average field strength in a large gap may be very much lower than the field strength immediately in front of the cathode.

over a wide range of field strengths (cf. Goodwin and Macfadyen 1953). Introducing a new variable

$$u = \int_0^x \eta_0 F^2 dx, \quad \dots\dots\dots (22)$$

and a constant (with respect to x)

$$u_d = \int_0^d \eta_0 F^2 dx, \quad \dots\dots\dots (23)$$

we have from (17)

$$\eta_0 F^3 dF = -\frac{4\pi I_c}{k_+} (\exp u_d - \exp u) du. \quad \dots\dots\dots (17a)$$

Integration and use of the boundary conditions yields

$$F^4 - F_c^4 = -\frac{16\pi I_c}{\eta_0 k_+} (u \exp u_d - \exp u + 1). \quad \dots\dots\dots (18a)$$

Equation (16) will again be given by (9) as

$$\left. \begin{aligned} I_c &= a F_c^2 \exp\left(-\frac{b}{F_c}\right), \\ a &= \frac{e^3}{8\pi\hbar\varphi}, \\ b &= \frac{4K\varphi^{3/2}}{3e}. \end{aligned} \right\} \quad \dots\dots\dots (16a)$$

where

and

Substituting in (18a) we find

$$F^4 - F_c^4 = -\frac{16\pi a F_c^2 \exp(-b/F_c)}{\eta_0 k_+} (u \exp u_d - \exp u + 1). \quad \dots\dots\dots (19a)$$

Owing to the complicated form of this equation a simple expression for F as a function of F_c cannot be obtained in general. However, in the vicinity of the cathode $u \ll u_d$, so that we have with the use of (22)

$$F^4 \simeq F_c^4 - \frac{16\pi a F_c^2}{k_+} \exp\left(-\frac{b}{F_c} + u_d\right) \int_0^x F^2 dx. \quad \dots\dots\dots (19b)$$

Assuming a trial solution for (19b) of the form

$$F = F_c \exp(-\gamma x), \quad \dots\dots\dots (24)$$

for sufficiently small values of x we find

$$\gamma = \frac{4\pi a}{k_+} \exp\left(-\frac{b}{F_c} + u_d\right). \quad \dots\dots\dots (25)$$

For larger values of x a very approximate solution gives $F \propto 1/x^{1/2}$, but, since the neglected effect of recombination will be becoming more important for larger distances from the cathode, it is reasonable to suppose that the field strength becomes practically constant for sufficiently large values of x . It is thus not

possible to derive an accurate expression for the apparent percentage decrease in breakdown strength, but sufficient calculations have been done to show whether the mechanism discussed could produce a measurable effect.

VI. DISCUSSION

(a) Solids

If any effect due to the cathode material is to be measurable, equation (8a) shows that not only must the field near the cathode be significantly lower than the breakdown field, but also that the field must be maintained at a lower value than breakdown for an appreciable distance across the specimen. Thus, in addition to the obvious condition $F_c^* < F^*$, we must have a condition on β (which measures the ability of the dielectric to maintain a reduced field). Three main cases can be distinguished for which (8a) gives

$$\left. \begin{aligned} D &\simeq 1 - \frac{F_c^*}{F^*}, & \text{if } \beta \ll F^*, \\ D &= \frac{F^* - F_c^*}{\beta} \left[1 - \exp \left(-\frac{\beta}{F^*} \right) \right], & \text{if } \beta \sim F^*, \\ D &\simeq 0, & \text{if } \beta \gg F^*. \end{aligned} \right\} \quad \dots \quad (8b)$$

It appears then that the conditions for a measurable effect of the cathode material reduce most simply to

$$F_c^* < F^* \quad \text{and} \quad \beta \lesssim F^*. \quad \dots \dots \dots (26)$$

Before endeavouring to decide whether these conditions can apply to real solids we rewrite (2a) and (15) substituting numerical values for universal constants, and in the case of (15) assuming that $W/\Delta V \sim 5$. This gives

$$1.55 \times 10^6 \frac{F_c^{*2}}{\varphi} \exp \left(-66 \frac{\varphi^{3/2}}{F_c^*} \right) = \sigma^* F^*, \quad \dots \dots \dots (2b)$$

and

$$\beta \sim 10^{-11} N d, \quad \dots \dots \dots (15a)$$

in which β , F_c^* , and F^* are in MV/cm, φ is in eV, σ^* in micromho cm⁻¹, N in electrons/c.c., and d in cm.

Working on polythene, Oakes (1948) gives $F^* \simeq 5$ with $d \simeq 5 \times 10^{-3}$, from which data using (15a) and (26) we find $N \lesssim 10^{14}$ if any such cathode effect is to be measurable. This seems a reasonable value for a substance such as polythene, and if $F_c^* < F^*$ the effect should be observable. Turning to (2b) we find that, on account of the extreme sensitivity of the exponential term to small changes in φ and F_c^* , a wide range of reasonable values can be covered. Thus the most that can be said in this case is that, for reasonable values of the parameters involved, the explanation given for the apparent decrease in the breakdown strength of polythene could be correct.

Taking the experiments of Cooper and co-workers (Calderwood, Cooper, and Wallace 1953; Cooper and Grossart 1953) on alkali halides we find $F^* \simeq 1$ for $d \simeq 5 \times 10^{-2}$. This would require $N \lesssim 2 \times 10^{12}$ if there is to be any measurable effect due to cathode material. This would appear to be very much too low a

value for the number of conduction and trapped electrons in an alkali halide at room temperature. It thus appears that an effect of the type discussed here would probably not be measurable in experiments on alkali halides—a conclusion which is in agreement with the findings of Cooper and his co-workers. It may be argued that the effect should be found in experiments on thinner specimens, but it seems from the figures that the thickness of the specimen would have to be reduced so much as to be impracticable. Furthermore, in the case of alkali halides, ionic currents and space charges would be expected to intervene, and their effect would be opposite to that of electronic currents and space charges in producing any cathode dependence of the breakdown strength.

(b) *Liquids*

From (24) the condition for a measurable effect of the cathode material will be

$$\text{or } \left. \begin{array}{l} \gamma^* x \sim 1, \text{ for } x < d, \\ \gamma^* d > 1, \end{array} \right\} \dots\dots\dots (27)$$

where γ^* is the value of γ when breakdown is about to commence at the cathode. Using (25) and (27) and substituting numerical values for universal constants we obtain

$$\frac{19.5 \times 10^{-6} d}{\phi k_+} \exp \left\{ -66 \frac{\phi^{3/2}}{F^{*}} + u_d^* \right\} > 1, \quad \dots\dots\dots (27a)$$

where u_d^* is the value of u_d when breakdown is about to commence at the cathode. (Note that equation (27a) may not be satisfied by indefinitely increasing d since then the neglect of recombination introduces serious error.)

Owing to the extreme sensitivity of (27a) to factors in the exponential term no definite predictions can be made but it is interesting to investigate orders of magnitude. Thus the various experimental workers previously quoted give for *n*-hexane: $F^* \sim 1$ MV/cm, $d \sim 5 \times 10^{-3}$ cm, $k_+ \sim 10^{-3}$ cm² V⁻¹ sec⁻¹, and $\eta \sim 10^4$ cm⁻¹ near breakdown. Substituting these values in (27a) we find that $\gamma^* d \gtrsim 1$ if $\phi \lesssim 1$ eV, which is a most reasonable value.

These calculations tend to support the opinion that the differences between the work of Lewis (1953a) (who found no measurable effect of cathode material) and that of Edwards (1951), Salvage (1951), and Goodwin and Macfadyen (1953) (who did find such an effect) may be well due to differences in the polish and cleanliness of the cathode surface. This seems to be so since the precise value of ϕ is very critical and any variation of it (such as that leading to emission from spots on the cathode) would completely alter the situation. In addition, some light is thrown on the experiments of Lewis (1953b) with a point-plane electrode system. With the point negative the non-uniformity of the field would be increased and the breakdown voltage decreased, while with the point positive the reverse would hold. This was in fact observed by Lewis, who concluded that the liquid was therefore stressed more highly when the plane was negative. However, this is not necessarily so, since space charge effects of the type discussed above would modify the maximum field strengths calculated by Lewis in such a way as to render them more nearly equal.

VII. REFERENCES

- CALDERWOOD, J. H., COOPER, R., and WALLACE, A. A. (1953).—*Instn. Elect. Engrs. Paper No.* 1457.
- COOPER, R., and GROSSART, D. T. (1953).—*Proc. Phys. Soc. Lond.* B **66** : 716.
- EDWARDS, W. D. (1951).—*Canad. J. Phys.* **29** : 310.
- FRÖHLICH, H. (1937).—*Proc. Roy. Soc. A* **160** : 230.
- FRÖHLICH, H. (1939).—*Proc. Roy. Soc. A* **172** : 94.
- FRÖHLICH, H. (1947).—*Proc. Roy. Soc. A* **188** : 521.
- GOODWIN, D., and MACFADYEN, K. (1953).—*Proc. Phys. Soc. Lond.* B **66** : 85.
- VON HIPPEL, A., and ALGER, R. S. (1949).—*Phys. Rev.* **76** : 127.
- LEWIS, T. J. (1953*a*).—*Instn. Elect. Engrs. Paper No.* 1488.
- LEWIS, T. J. (1953*b*).—*Proc. Phys. Soc. Lond.* B **66** : 425.
- LOEB, L. B. (1939).—“Fundamental Processes of Electrical Discharge in Gases.” (John Wiley & Sons: New York.)
- MOTT, N. F., and GURNEY, R. W. (1949).—“Electronic Processes in Ionic Crystals.” (Oxford Univ. Press.)
- MOTT, N. F., and SNEDDON, I. N. (1949).—“Wave Mechanics and its Applications.” (Oxford Univ. Press.)
- OAKES, W. G. (1948).—*J. Instn. Elect. Engrs.* I **95** : 36.
- SALVAGE, B. (1951).—*Proc. Instn. Elect. Engrs.* IV **98** : 15.

THE RESIDUAL RANGE OF DELAYED PARTICLES IN EXTENSIVE AIR SHOWERS

By V. C. OFFICER* and P. J. ECCLES*

[Manuscript received May 28, 1954]

Summary

The arrival times of the penetrating particles of extensive air showers relative to that of the electrons have been studied by means of short reaction time Geiger counters. A 50 channel hodoscope has been used to identify the penetrating particles and measure their residual ranges where these lay between 15 and 30 cm of lead. From observations on 782 showers of median density 70 particles m^{-2} in which a penetrating component was detected, it was deduced that between 3 and 9 particles in 10^4 shower particles from showers of median density 28 particles m^{-2} , have delays $< 5 \times 10^{-8}$ sec and are able to penetrate 15 cm of lead. It follows that 81-94 per cent. of the delayed particles found by Jelley and Whitehouse (1953) without the use of absorbers, must be stopped by 15 cm of lead. This indicates a height of production below 0.8 km for at least half of these particles.

Two of the penetrating delayed μ -mesons were stopped in the lead, and their heights of production calculated to be $1.0^{+0.7}_{-0.4}$ km and $4.7^{+4.5}_{-2.2}$ km on the assumption that the delays were due to velocity differences. Thirty other μ -mesons for which the individual time lags were not significant were also stopped in the lead and gave a mean delay indicating production below an altitude of 250 m. The remaining 208 μ -mesons which did not stop could not be assigned a height of production.

One delayed proton was found in the total of 29 delayed events observed, and nine events could have been oblique particles lagging on the electrons by virtue of path differences.

I. INTRODUCTION

Some particles in extensive air showers can arrive at the plane of observation later than others if they have travelled from high in the atmosphere with a slightly lower velocity, or if their path lengths have been different. It is to be expected that the heavier shower particles, the mesons and nucleons, will have the largest delays. If a delayed particle is identified and its residual range measured as well as its delay with respect to the shower electrons, its height of production can be found provided the delay was due to a velocity difference. In this way it should be possible to find the distribution of the heights of production and energies at production for the low energy μ -mesons in air showers. If, in a simple analysis, the fact that mesons traverse the early part of their path as π or heavier mesons is neglected, an overestimate of the height of production will result.

Several previous investigations have been made into the delayed particles in air showers. McCusker, Ritson, and Nevin (1950) found no particles delayed

* Physics Department, University of Melbourne.

by more than 150 μsec (i.e. 150×10^{-8} sec). Mezzetti, Pancini, and Stoppini (1951) found penetrating particles with delays between 10 and 120 μsec . The mean delay of particles able to penetrate 10 cm of lead was found by Officer (1951) to have an upper limit of 2 μsec . This latter work has now been continued after developing short reaction time Geiger counters. Recently Jelley and Whitehouse (1953) using a large unshielded liquid scintillation counter have found 0.6 per cent. of shower particles to have delays between 3 and 70 μsec . The distribution of delays could be represented by an exponential function with half the delays less than 10 ± 2 μsec . Three large liquid scintillation counters have recently been used by Bassi, Clark, and Rossi (1953) to study the distribution of arrival times of both the electrons and the penetrating particles at sea-level. They found that at a given instant most of the electrons lie in a flat disk between 1 and 2 m thick, and the particles able to penetrate 20 cm of lead follow less than 3 m behind in a disk between 2 and 3 m thick. With widely spaced counters they studied the angular distribution of shower axes and the curvature of shower fronts. In the present work time lags arising from the inclination of showers or curvature of shower fronts are not important, since the counter tray spacing was only 2 m.

II. APPARATUS AND PROCEDURE

The showers were detected by a set of three Geiger counter trays arranged at the apices of a 2 m equilateral triangle and connected in threefold coincidence with a resolving time of approximately 1 μsec . Two of the trays were composed of short reaction time counters and connected to the two channels of the time interval measuring apparatus. One of the timing trays was shielded with 15 cm of lead and the other was unshielded. The third tray contained ordinary counters and supplied pulses to the coincidence circuit only. Under the shielded tray were layers of lead and hodoscope counters for residual range measurement and penetrating event identification. A threefold coincidence caused the time interval between the pulses from the two timing trays to be recorded at the same time as the configuration of hodoscope counters discharged.

The shielded timing tray, 359 cm^2 in area, consisted of 26 short reaction time counters with nickel cathodes 4.6 mm in diameter (Officer and Eccles 1954). The unshielded timing tray, 450 cm^2 in area, consisted of 30 short reaction time Maze-type counters with glass cathodes 5 mm in inside diameter. The unshielded third tray was 583 cm^2 in area, and consisted of 10 conventional counters 2 cm in diameter.

The small glass counters had anode wires of tungsten 0.004 in. in diameter mounted accurately on the axis of soft glass tubes of 5 mm bore. A layer of "Aquadag" painted on the outside of the tubes and extending over 30 cm of their length was connected to the negative voltage supply. Earthed guard rings of "Aquadag" 1 cm from the ends of the cathode coating prevented electrical leakage from the negative supply over the surface of the glass to the signal circuits connected to the anode wires. A coat of clear "Glyptal" protected the insulating glass surfaces and the "Aquadag". Six of these counters were mounted in a plane and sealed to the same quenching mixture

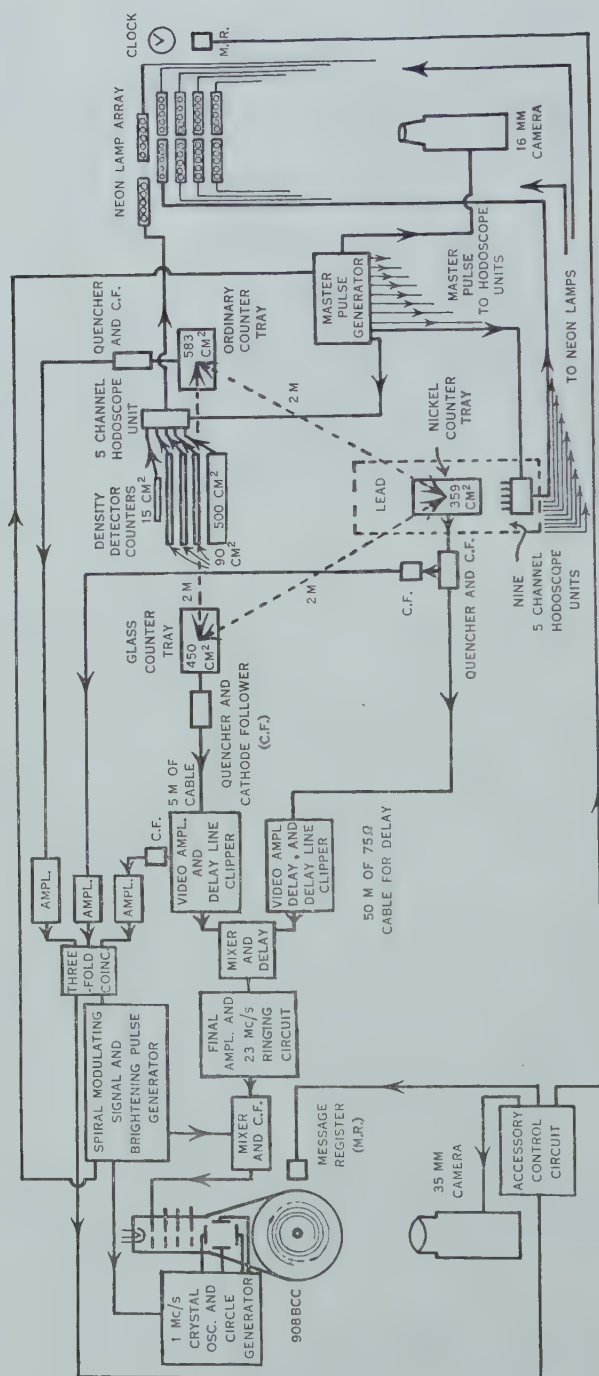


Fig. 1.—Block diagram of circuits and plan of counter arrangement.

reservoir, a soft glass cylinder 4 cm in diameter and 43 cm long. The processing included baking at 200 °C under vacuum. The filling was 22.5 cm of 9 : 1 argon-ethyl formate mixture for the same starting and working voltages, 970 and 1120 V, as for the nickel counters. All the counters in each tray were connected in parallel and provided with a single external quenching circuit per tray to prolong their life. Direct coupling to the quenching circuit made the anode wires 300 V above ground potential.

The performance of the glass counters was similar to that of the nickel counters, although, as will be shown in Section 3 (a), the end effects were greater. The relative reaction time distribution for the two full trays, one of glass and one

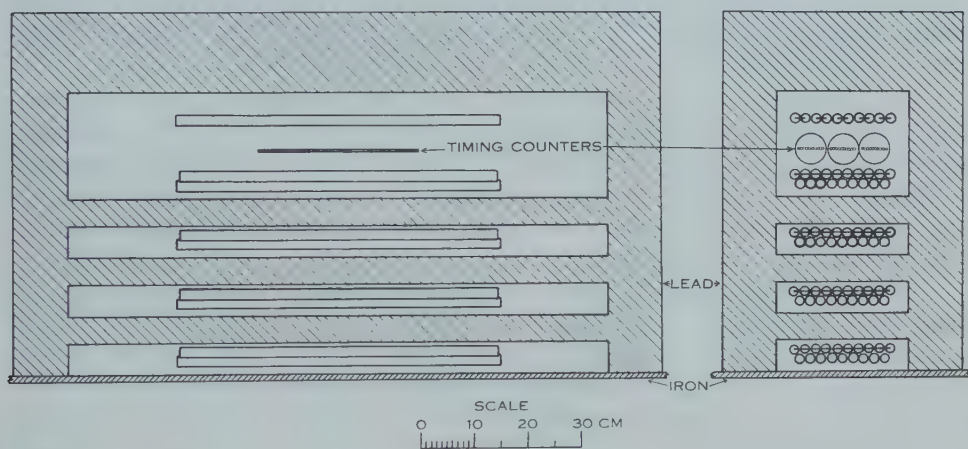


Fig. 2.—Counter and lead arrangement for residual range measurement, penetrating event identification, and timing.

of nickel counters, had only slightly greater spread than that reported by Officer and Eccles (1954) for much smaller trays of nickel counters operated in parallel. The standard deviation of the distribution was 1.6 ± 0.06 cμsec.

The two short reaction time trays were connected to the timing channels as shown in the block diagram of Figure 1. The timing apparatus was similar to that described earlier by Officer (1951), but brilliance modulation instead of radial deflexion was used to put the timing information on the spiral time base as outlined previously (Officer and Eccles 1954).

Figure 2 shows the arrangement of the hodoscope counters around the nickel timing tray, and the disposition of the lead for shielding and residual range measurement. Counters joined by horizontal lines are connected to the same hodoscope channel. The rest have one hodoscope channel each. The individually connected counters gave information about the nature of the penetrating event, and the guard layers made the determination of residual range more certain. A hodoscope resolving time of 20 μsec was used. Residual range measurements were confined to the 15–30 cm of lead interval, but calcula-

tion showed that mesons having these residual ranges and originating higher than a few kilometres above the apparatus, should have measurable time lags. A thickness of 15 cm of lead was chosen as the top shield, rather than the 20 cm thickness usually used to isolate the penetrating component of air showers. The aim was to detect as many as possible short residual range, possibly highly delayed, particles. The hodoscope was relied upon to reveal the cases in which these particles were accompanied by other phenomena.

A simple shower density detector consisting of five counters connected to hodoscope channels was placed as shown in Figure 1. Its chief purpose was to enable an estimate of the number of hits scored on the unshielded timing tray to be made. This allowed recognition of cases in which an apparently late pulse from the shielded tray could be due to the selection of the shortest of several counter reaction times from the unshielded tray, rather than to a genuinely delayed single penetrating particle striking the shielded tray.

The photographs of the cathode-ray tube trace and those of the hodoscope information displayed on neon lamps were correlated by means of message registers appearing in each photograph. The time intervals were read from the photographs of the spiral trace by projecting them on to a vernier reading circle, and measuring between the sharp trailing edges of the bright dashes produced by the timing pulses.

The timing pulses were separated by an artificial delay to avoid overlapping when short intervals were measured, and they were both delayed sufficiently to enable time lags up to 30 ± 10 μ sec to be measured. The fluctuation in the length of time base available for measurement was due to the spread in the reaction times of the ordinary counters in the third tray. The time origin was found by placing the two unshielded trays on top of the lead shielding of the nickel tray, and recording the pulses produced by penetrating particles of the normal cosmic ray flux passing through the trays. Three to four hundred observations were required, and a time of flight correction was applied. The origin was determined in this way at intervals throughout the run, and a histogram containing the whole 2064 of these observations is shown in Figure 3 (a).

During some of the origin tests the particles selected by the counter telescope were used to test the performance of the hodoscope. The main performance monitoring of the whole equipment was a daily routine of waveform checking. Since the spiral time base was generated from the output of a 1 Mc/s quartz crystal oscillator no time base calibration was required.

III. RESULTS

(a) *The Abundance of Delayed Events*

A total of 782 showers was recorded in 1543 hr. The time interval histogram for all these showers, except a few in which a cathode-ray "ghost image" obscured the record, is shown in Figure 3 (b). The histogram of Figure 3 (a) contains all the time origin observations taken with single mesons passing through both timing trays. It is normalized to the same area as that of the shower histogram and its mean, after a 10^{-9} sec time of flight correction, is taken as the time origin for the shower histogram.

The time intervals referred to this origin are the differences between the times of detection of the pulses from the two timing trays. Positive intervals indicate that the pulse from the shielded tray was the later, and negative intervals indicate that the unshielded tray gave the later pulse.

It can be seen that the shower histogram has a greater spread than the origin histogram, especially at the base. If delays greater than 5 μsec are considered significant, 29 showers had significantly delayed penetrating particles. If exceeding 5.4 μsec , the extreme upper limit of the origin observations, is taken as the criterion of significance there are 23 significantly delayed events.

There is also an extension of the shower histogram in the negative direction, but there is a similar tail on the origin distribution although it has a smaller fraction of the observations in it. The extra negative lags could not be due to a genuinely delayed particle discharging the unshielded tray and a prompt penetrat-

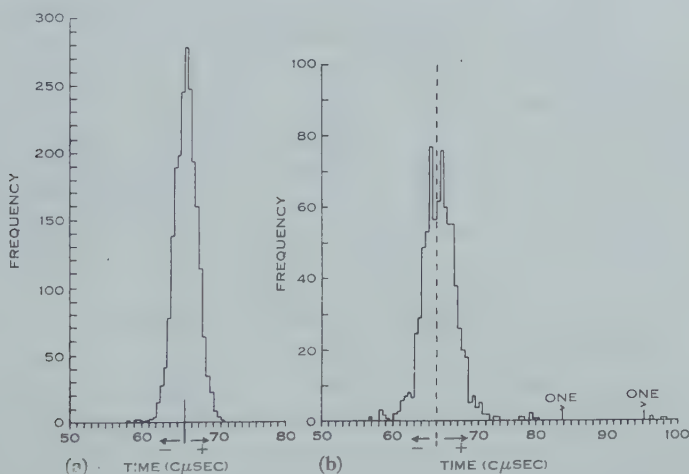


Fig. 3.—Time origin and total shower time interval histograms. (a) Origin histogram, 2064 observations; (b) all showers histogram, 776 showers (one 24.9 μsec interval and two cases with one pulse only).

ing one striking the shielded tray, if the total abundance of delayed particles is as found by Jelley and Whitehouse. Only one negative lag would be expected from this cause. By selecting with a counter telescope a beam of penetrating mesons passing through the ends of the counters in the unshielded tray, it was shown that the negative tail could be due to an end effect in the small diameter glass counters. The counter telescope arrangement used for the time origin observations gave a smaller probability for particles passing through the counter ends than did the shower arrangement. A similar test for end effect in the nickel counters of the shielded tray showed that the effect was less marked for these counters. The positive tail of the test histogram exhibited an almost linear decrease from 4 μsec to an upper limit of 7.3 μsec , whereas the negative tail found in the test of the glass counters had a similar shape to that in Figure

3 (a) and extended as far as $-15.3 \mu\text{sec}$ with no indication that a limit had been reached.

When the different end structures of the two counter types are considered it can be seen that the great difference in behaviour is reasonable. The sensitive length of the nickel counters is terminated by the 0.002 in. diameter anode wire entering a "Monel" capillary tube 1.7 mm in outside diameter with a hemispherical end. The nickel cathode cylinder is 4.6 mm in inside diameter. In the glass counters the sensitive length terminates more slowly. The 0.004 in. diameter anode wire emerges from the "Aquadag" coated cathode section of the 5 mm bore glass tube and traverses a 1 cm uncoated length before passing through an earthed guard ring. The distribution of potential over the uncoated glass governs the form of the electric field at the ends. The larger capacity of the glass counters, resulting from the use of thicker anode wires, may also have an influence on the importance of end effects, and different discharge propagation velocities could result in different lengths of end section being affected.

If exceeding $7.3 \mu\text{sec}$ is taken as the criterion for significance of delayed events, the result should be a safe lower limit to their abundance, as end effect delays of $7.3 \mu\text{sec}$ are rare even when a beam passing through the end section is selected. In fact, as will be shown in Section III (b), much of the positive tail below $7.3 \mu\text{sec}$ is due to multiple hits on the unshielded tray. We are left with a lower limit of 11 significantly delayed events. The upper limit remains at 29. Then from the expression for the rate of detection of showers containing significantly delayed penetrating particles,

$$\int_0^\infty K \Delta^{-\gamma} (1 - e^{-S_1 \Delta}) (1 - e^{-S_2 \Delta \epsilon}) (1 - e^{-S_3 \Delta}) d\Delta \quad \text{hr}^{-1},$$

where $K=787$, $\gamma=2.4$, and S_1 , S_2 , S_3 are the tray areas, the fraction ϵ of shower particles significantly delayed and able to penetrate 15 cm of lead lies between 3 and 9 in 10^4 . The values of K and γ are those given by Singer (1951). This calculation assumes that ϵ is independent of the shower density Δ . The calculated median density of the penetrating extensive showers detected is 70 particles m^{-2} , but it could be considered that the showers that would discharge the three trays if there was no shield were being searched for penetrating delayed particles. The calculated median density of these showers is 28 particles m^{-2} .

The complete shower histogram contains one large negative lag of $-41.1 \mu\text{sec}$ and two cases in which only the pulse from the shielded tray appeared on the trace. It is very likely these are due to late quenching circuit pulses from the unshielded tray. With only one quenching circuit to a whole tray of counters an occasional counter discharge will occur before the quenching circuit has fully recovered from the previous one, and therefore give rise to a delayed quenching pulse. Normally the quenching pulse arrives within the clipping time of the delay line clipper and is not resolved from the counter pulse, but a late quenching pulse should appear as a third pulse on the trace. However, since the coincidence circuit requires the large amplitude of the quenching pulse, the time base starts late on these occasions and only one or two pulses may be seen. Since the counting rate, 7.1 sec^{-1} , in the heavily shielded

nickel tray was low, and that of the glass tray, 44.9 sec^{-1} , very high, the effect was confined to the latter tray. Radioactive potassium in the soft glass (A. G. Fenton, personal communication), as well as lack of shielding, was responsible for the high rate. About 1 in 200 quenching pulses from the glass tray was observed to be late. It follows from this and the counting rates that 1 in 1240 from the nickel tray would be late, and since there were only 782 showers observed the effect could not make a significant contribution to the delayed events.

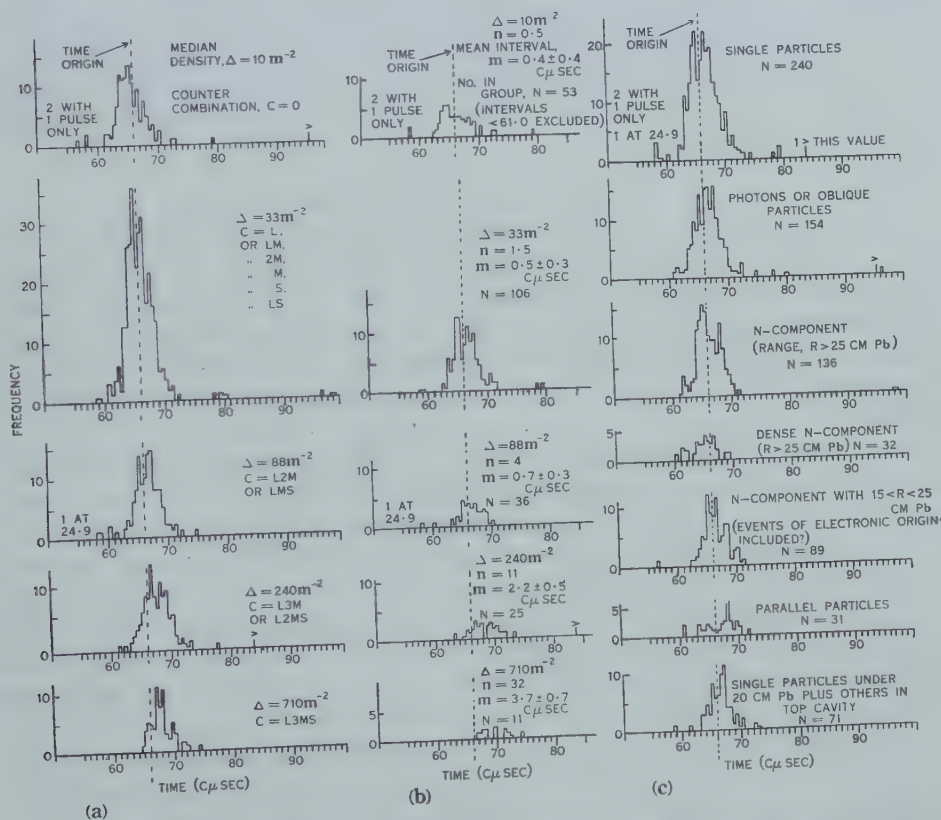


Fig. 4.—Analysis of time intervals according to shower density and nature of penetrating event. (a) Density analysis; (b) density analysis for single penetrating particle events; (c) nature of penetrating event analysis. n is the average number of hits on an area equal to that of the unshielded tray in showers of density Δ . In (c) there were 17 observations in a miscellaneous group.

Accidental coincidences were also insignificant. Calculation shows that the probability that there are no accidentals in the 782 showers is 0.6.

(b) The Influence of Shower Density

Figure 4 (a) shows an analysis of the shower results according to the density information supplied by the density detector group of counters. The calculated median shower density is quoted for each histogram and also the combinations of large, medium, and small density detector counters discharged. The median

number of hits scored on the unshielded tray by showers of the stated median densities, detected or not, are also quoted. A progressive shift in the positive direction of the histogram mean values as the shower density increases is clearly shown. This effect is mainly responsible for the broadening of the total shower histogram of Figure 3 (*b*). The effect arises from the fact that in showers of high density the unshielded timing tray is hit by many more particles than the shielded tray which often receives only one hit. This means that the shortest of many counter reaction times is chosen from the unshielded tray, giving an abnormally short reaction time and making the pulse from the shielded tray appear to be late. This effect has been noticed earlier by Officer (1951) and has been discussed more fully by Officer and Eccles (1954) for the case of simultaneous hits. Here of course any spread in the arrival times of the particles will modify the effect.

(*c*) *The Nature of the Delayed Events*

In Figure 4 (*c*) the time intervals are analysed according to the information on the nature of the penetrating events given by the hodoscope. Most of the positive intervals greater than 5 μ sec are found in the single particle group. This group consists of cases in which a single penetrating particle hit the shielded timing tray and continued into the lead layers below. From the rarity of interactions produced in the lead by these particles they were identified as μ -mesons.

The positive intervals greater than 7.3 μ sec are divided about equally between the single particle group and the very oblique particle or photon group. In the latter group either no hodoscope counter at all or just one in the top lead cavity was discharged. These events could equally well be explained as long range photons from burnt out cascades in the 15 cm of lead above (Greisen 1949) or very oblique particles. The photons should predominate, but it is possible that the significant time lags were produced by oblique particles that arrived late through having traversed a large component of path perpendicular to the shower axis.

There was one significantly late event in the *N*-component group. In this event a particle with range greater than 30 cm of lead produced secondary particles that could penetrate 10 cm of lead. If the particle was a μ -meson it would have required an energy of at least several kMeV, and could not have travelled sufficiently slowly to lag 32 μ sec on the shower electrons even if it came from the top of the atmosphere. It was considered likely to be a proton which could have arrived from an altitude of 20 km with this lag and a residual energy of about 7 kMeV.

It will be noticed that the very dense *N*-component group, in which most of the shielded counters were fired even under 30 cm of lead, shows an excess of negative lags. This is a density effect of opposite sign to the previous one because the shielded tray frequently received many more hits than the unshielded one.

Figure 4 (*b*) shows the single particle group analysed according to shower density. As before the upward shift in the mean values with increasing shower

density is clear. It will be noticed, however, that there are both large and small, possibly significant positive lags in the low density groups where multiple hits on the unshielded tray would be very unlikely.

(d) *Residual Ranges and Heights of Production*

The delays of μ -mesons relative to particles travelling with virtually the velocity of light have been calculated for various residual ranges in lead at sea-level and common starting heights. These delays, shown in Figure 5, could apply to delays of μ -mesons relative to the shower electrons if it is assumed that the nuclear cascade, which gives rise to both the electrons and the mesons, has kept pace with the electrons down to the point of formation of the mesons

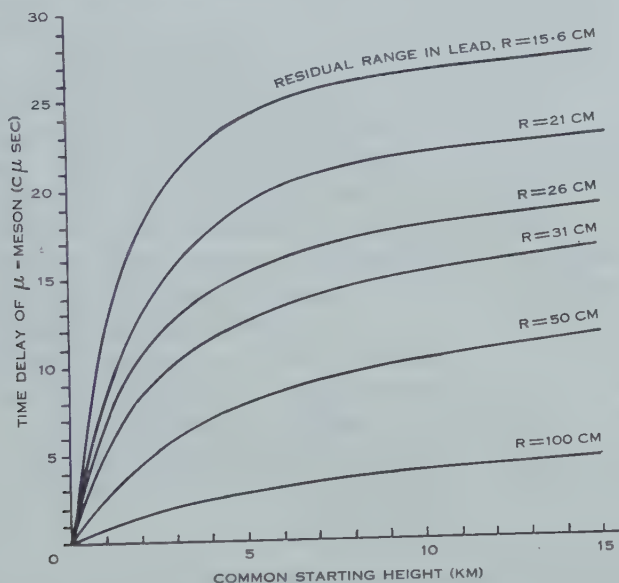


Fig. 5.— μ -Meson delays with respect to a particle having virtually the speed of light plotted against their common starting height for various residual ranges in lead.

in question. The fact that the mesons will have traversed the early part of their path as π or heavier mesons has been neglected. These simplifications could lead to an overestimate of the height of production of a meson with given delay and residual range.

Of the particles with significant time lags only two were brought to rest in the lead below the shielded tray. Both had ranges between 25 and 30 cm of lead and both were considered to be μ -mesons. The first had a delay of 8.4 ± 1.6 $c\mu$ sec where the error quoted is the standard deviation of the time origin distribution, but the delay should be reduced by about 2.5 $c\mu$ sec as it came from the 710 particles m^{-2} shower density group. With as many as 32 hits on the unshielded tray its reaction time should be very small, and the time origin distribution under these conditions should be roughly rectangular extending from 0 to about 5 $c\mu$ sec, the approximate extreme value of the normal origin

distribution (Officer and Eccles 1954). With such a large number of hits the time required to collect sufficient charge for detection may be reduced as well as the electron drift time, making a somewhat larger correction necessary. If the lag is taken as 5.9 ± 1.6 μsec the height of production would be $1.0^{+0.7}_{-0.4}$ km above the apparatus, and the extreme range of heights 0.1–3.6 km if ± 5 μsec is taken as the extreme timing error.

The second particle came from a low density group and its lag of 13.4 ± 1.6 μsec needed no correction. Its height of production was $4.7^{+4.5}_{-2.2}$ km with extreme limits of 1.4–13 km.

There were 30 single particles without significant lags stopped* in the lead, their residual ranges lying between 15 and 30 cm. They were all considered to be μ -mesons. Eighteen of them came from low density groups and gave a mean lag of $0.4 \pm 0.5_7$ μsec . It should be safe to say that the lag was less than 1.5 μsec and their mean height of production therefore less than 250 m.

The mean lag for the whole single particle or μ -meson group analysed in Figure 4 (b) is $0.8_8 \pm 0.1_7$ μsec . The negative lags in excess of -5 μsec were omitted, but if the positive lags greater than 5 μsec are also omitted there is still a positive mean lag of $0.5_3 \pm 0.1_4$ μsec . This positive value could easily be due to the influence of the density effect, but the group with median density 10 particles m^{-2} should be free from this uncertainty. The mean value of this lowest density group is $0.4_5 \pm 0.4$ μsec with the large negative lags omitted and $-0.0_6 \pm 0.3$ μsec with the large positive lags also omitted. The fact that the standard deviation of this group, 3.0 ± 0.3 μsec is significantly greater than 1.6 ± 0.06 μsec for the origin distribution may be related to the thickness and spacing of the electron and meson shower disks found by Bassi, Clark, and Rossi (1953).

IV. DISCUSSION

The present results show that between 3 and 9 in 10^4 air shower particles are significantly delayed and are also capable of penetrating at least 15 cm of lead. "Significantly delayed" could be interpreted to mean that the particles have delays greater than 5 μsec on the average, provided the abundance of delayed particles does not vary too rapidly with delay in the region 0–10 μsec , the approximate extreme width of the time origin distribution. From the results obtained by Jelley and Whitehouse (1953) without absorbers, 47 in 10^4 shower particles are delayed more than 5 μsec with respect to the electrons. This means that between 81 and 94 per cent. of the delayed particles in air showers are absorbed by 15 cm of lead. The showers detected by Jelley and Whitehouse had about 1/3 of the density of the present ones, but they found no correlation between delay and density.

* In the group of 32 particles recorded as stopped in the lead six could have been erroneously included through their having escaped from the ends of the hodoscope array. This was calculated from the zenith angle distribution of penetrating shower particles obtained from the hodoscope records, and to be published later together with other data on the penetrating component of air showers.

The upper limit to the residual range of the bulk of the delayed particles thus obtained allows something to be learned about their height of production. If the particles are μ -mesons and the delays due to velocity differences, at least half of them must have been produced below an altitude of 0.8 km, since Jelley and Whitehouse found that half of them had delays less than 10 μ sec. Similarly at least three-quarters of them were produced below 2.7 km, but the present upper limit to the residual range would permit production of the highly delayed particles at almost any height. If the delays are due to oblique particles having very different path lengths from the electrons, a low height of origin would still be indicated. If the delayed particles are protons rather than μ -mesons the heights of production would be very much lower, but it seems likely that Jelley and Whitehouse would have obtained larger scintillation pulses had the particles been protons with residual range less than 15 cm of lead.

The present results for μ -mesons with residual ranges between 15 and 30 cm of lead also lead to low heights of origin, less than 250 m on the average. Little can be said about the heights of production of the 208 μ -mesons that did not stop in the lead. The mean delay for the whole μ -meson group 0.88 ± 0.17 μ sec agrees with that found by Bassi, Clark, and Rossi (1953), but this may be a coincidence as their result should be free from the influence of multiple hit effects which are important here when dense showers are included. The mean of the 10 particle m^{-2} shower density group should be free from multiple hit effects, but it lacks precision and could be consistent with Bassi, Clark, and Rossi's result.

Since Jelley and Whitehouse find 0.6 per cent. of shower particles with delays between 3 and 70 μ sec, and it is known that about 1 per cent. of shower particles are penetrating μ -mesons, it appears, when their results are combined with the present results, that the production of low energy mesons in air showers at low altitudes is about as great as the production of high residual energy ones at all altitudes. This is equivalent to an attempt at deducing the vertical distribution of meson production from measurements on residual ranges and delays alone. It would be valid only if all mesons travelled parallel to the shower axis and did not decay. In practice mesons produced at high altitude with low enough energy to have a large time lag and small residual range at sea-level, may diffuse towards the outer extremities of the shower or decay before reaching sea-level. Then an apparatus such as the present one, which is biased towards detecting showers near the core, might be expected to receive most of its low energy mesons from low altitudes. A false impression of the production at high altitudes would be obtained since only the high energy mesons would be received from those altitudes. It therefore seems desirable to locate the shower core in any such experiment. It might be found that the highly delayed particles become more penetrating as the distance from the core increases provided sufficient of those produced at high altitude survive.

V. CONCLUSIONS

In extensive air showers of median density 28 particles m^{-2} at sea-level, between 3 and 9 in 10^4 shower particles are delayed by more than 5 μ sec with

respect to the electrons, and are able to penetrate at least 15 cm of lead. When this is combined with the results obtained by Jelley and Whitehouse it appears that 81–94 per cent. of the delayed particles in air showers are absorbed by 15 cm of lead. As the delayed particles are likely to be mostly μ -mesons, it follows that at least half of the delayed particles detected without absorbers were produced below an altitude of 0.8 km.

In extensive air showers of median density 28 particles m^{-2} , μ -mesons having residual ranges between 15 and 30 cm of lead have a mean height of production below 250 m.

In a total of 782 observed showers two detected μ -mesons were both significantly delayed and brought to rest in 25–30 cm of lead. Their most probable heights of production were calculated to be $1.0^{+0.7}_{-0.4}$ km and $4.7^{+4.5}_{-2.2}$ km.

One penetrating delayed proton was observed but it is likely that most particles in air showers delayed by more than a few μsec are μ -mesons.

It is likely that path differences rather than velocity differences caused some of the delays observed.

VI. ACKNOWLEDGMENTS

It is a pleasure to thank Professor L. H. Martin for his continued interest and support. Part of the cost was met from funds provided by C.S.I.R.O.

VII. REFERENCES

- BASSI, P., CLARK, G., and ROSSI, B. (1953).—*Phys. Rev.* **92**: 441.
GREISEN, K. (1949).—*Phys. Rev.* **75**: 1071.
JELLEY, J. V., and WHITEHOUSE, W. J. (1953).—*Proc. Phys. Soc. Lond. A* **66**: 454.
McCUSKER, C. B. A., RITSON, D. M., and NEVIN, T. E. (1950).—*Nature* **166**: 400.
MEZZETTI, L., PANCINI, E., and STOPPINI, G. (1951).—*Phys. Rev.* **81**: 629.
OFFICER, V. C. (1951).—*Aust. J. Sci. Res. A* **4**: 526.
OFFICER, V. C., and ECCLES, P. J. (1954).—*Proc. Phys. Soc. Lond.* (in press).
SINGER, S. F. (1951).—*Phys. Rev.* **81**: 579.

MOMENTUM DISTRIBUTION AND CHARGE RATIO OF μ -MESONS AT ZENITH ANGLES IN THE EAST-WEST PLANE

By J. R. MORONEY* and J. K. PARRY*

[Manuscript received June 2, 1954]

Summary

The momentum distribution and charge ratio of the penetrating component of the cosmic radiation at sea-level have been determined over the momentum range 0.24–58 BeV/c at a geomagnetic latitude of 47°S. The measurements were performed in the vertical direction and at zenith angles of 30° and 60° in the eastern and western azimuths.

An attempt has been made to calculate the sea-level spectra at these zenith angles on the basis of a simplified continuous production process. A comparison with the measurements indicates that, although the calculations describe the general behaviour, the quantitative agreement is unsatisfactory. The charge ratios measured at zenith angles in the western azimuth increase, and those in the eastern azimuth decrease, as the momentum decreases. This is explained as a secondary effect due to curvature of the meson trajectories in the magnetic field of the Earth. Information has been obtained on the dependence of the exponent n on the momentum, where n is defined by the intensity-zenith angle relation, $I_\theta = I_0 \cos^n \theta$. From the value 3.3 at 0.3 BeV/c, n approaches zero at high momentum. The radiation is approximately isotropic above 20 BeV/c.

I. INTRODUCTION

In recent years a number of careful studies (Wilson 1946 ; Caro, Parry, and Rathgeber 1951 ; Owen and Wilson 1951 ; Beretta *et al.* 1953 ; and others) have been made of the momentum distribution and charge ratio of the penetrating component of cosmic radiation incident in the vertical direction. However, no comparable information is available on the radiation observed at inclinations to the vertical.

In an endeavour to extend the experimental knowledge in this region a measurement has been made of the momentum distribution and charge ratio of the penetrating component at angles of 30° and 60° in the east-west plane. The experiment was performed at sea-level at a geomagnetic latitude (Melbourne) of 47°S.

II. METHOD

The spectrometer used to determine the momentum and sign of single charged particles has been described previously (Caro, Parry, and Rathgeber 1951). The particles are deflected in the air gap of an electromagnet and the extent of the deflexion is measured by trays of Geiger counters. Additional counters under 10 cm of lead identify the penetrating component. It should be emphasized that this technique discriminates against penetrating particles

* Physics Department, University of Melbourne.

accompanied by shower particles; an unambiguous interpretation is possible only for single particles.

In order to cover the maximum momentum range, from the momentum cut-off imposed by the 10 cm of lead to the resolving limit of the instrument, the measurements are conducted at two values of the magnetic field. The momentum ranges covered at 1900 and 13,500 G are respectively 0.24–10 BeV/c and 1.3–70 BeV/c. At low momenta the measured spectrum is subject to instrumental distortion due to a magnetic cut-off effect. The results are corrected for this effect as described in the previous paper.

III. RESULTS

The results have been obtained in 10 separate measurements; those taken at 13,500 G in the vertical direction have been published previously (Caro, Parry, and Rathgeber 1951) but are included here for completeness. Table 1 gives the number of particles analysed in each measurement. The results of each determination were analysed using the methods discussed in the previous

TABLE 1
DETAILS OF THE TOTAL NUMBER OF PARTICLES RECORDED IN EACH MEASUREMENT AND THE ABSOLUTE INTENSITY AT EACH ZENITH ANGLE

Measurement	Number of Particles Recorded		Corrected Intensity (sterad ⁻¹ cm ⁻² sec ⁻¹)
	1900 G	13,500 G	
Vertical	3971	2127	$(8.20 \pm 0.26) \times 10^{-3}$
30 °E.	1304	2797	$(6.22 \pm 0.27) \times 10^{-3}$
30 °W.	1523	1410	$(6.12 \pm 0.24) \times 10^{-3}$
60 °E.	1230	1234	$(1.90 \pm 0.09) \times 10^{-3}$
60 °W.	2168	1385	$(1.83 \pm 0.09) \times 10^{-3}$

paper, corrections for magnetic cut-off were applied, and the differential momentum spectra constructed. The particles were grouped into momentum intervals chosen to give adequate momentum resolution with good statistics. The weighted mean momentum of each interval was used in plotting the results. Each 13,500 G determination of the momentum spectrum was normalized to its 1900 G counterpart by equating the number of particles recorded with momentum greater than 1.3 BeV/c.

Since the instrument was inoperative for 20 sec while each record was produced it has been necessary to correct the counting rate for this dead time. The absolute intensity at each zenith angle was computed from the known geometry of the counter telescope. These intensities are shown in Table 1.

The differential momentum spectra of the positive particles, the negative particles, and the total penetrating component normalized to the correct absolute intensities are plotted in Figures 1 (a)–(e) for each zenith angle. The smooth curves are drawn through by eye and are intended to represent the best fit to the results.

IV. DISCUSSION

The penetrating component at sea-level is known to consist predominantly of μ -mesons. An estimate based on the work of Mylroi and Wilson (1951) indicates that less than 0.5 per cent. of the penetrating particles recorded in the

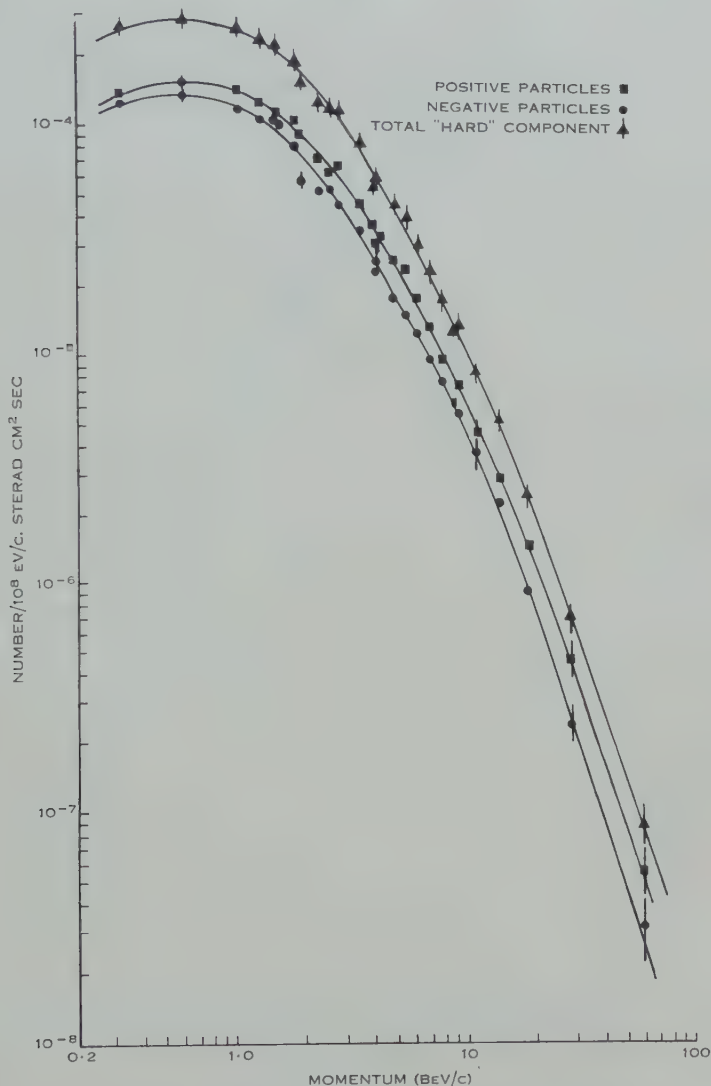


Fig. 1 (a).—Vertical differential momentum spectra.

vertical direction are protons. Other possible contributions are from high energy single electrons and π -mesons. Since only events in which a single counter under the lead is discharged are accepted for analysis, the contribution from high energy electrons may be neglected. Little is known of the intensity of high energy single π -mesons at sea-level but there is no evidence to suggest that this

is appreciable. Since no calculation in the following section is accurate to better than 5 per cent. it is felt that no significant error is made if, for the purpose of comparison with calculations, the penetrating component in all measurements is identified with the μ -meson component.

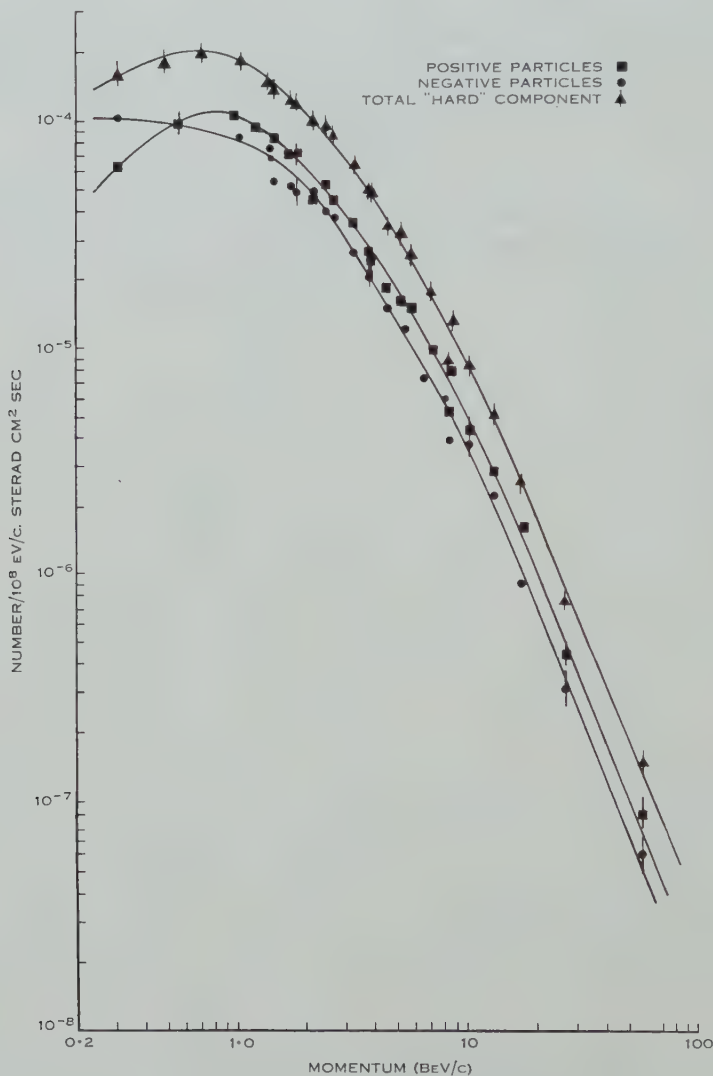


Fig. 1 (b).—30 °E. differential momentum spectra.

(a) *The Total Spectra at Zenith Angles of 0°, 30°, and 60°*

(i) *Introduction.*—The smooth curves representing the differential spectra of the μ -meson component in the five directions are shown together in Figure 2. It is apparent that the main effect observed is the increasing attenuation of the low momentum intensity with zenith angle. The differences between the spectra

measured in the eastern and western azimuths for a given zenith angle, and between all spectra above 20 BeV/c, are not considered significant when the statistical accuracy of the points is taken into account. Although at the latitude of this experiment a small east-west asymmetry is known to exist, 0.01 at 30°

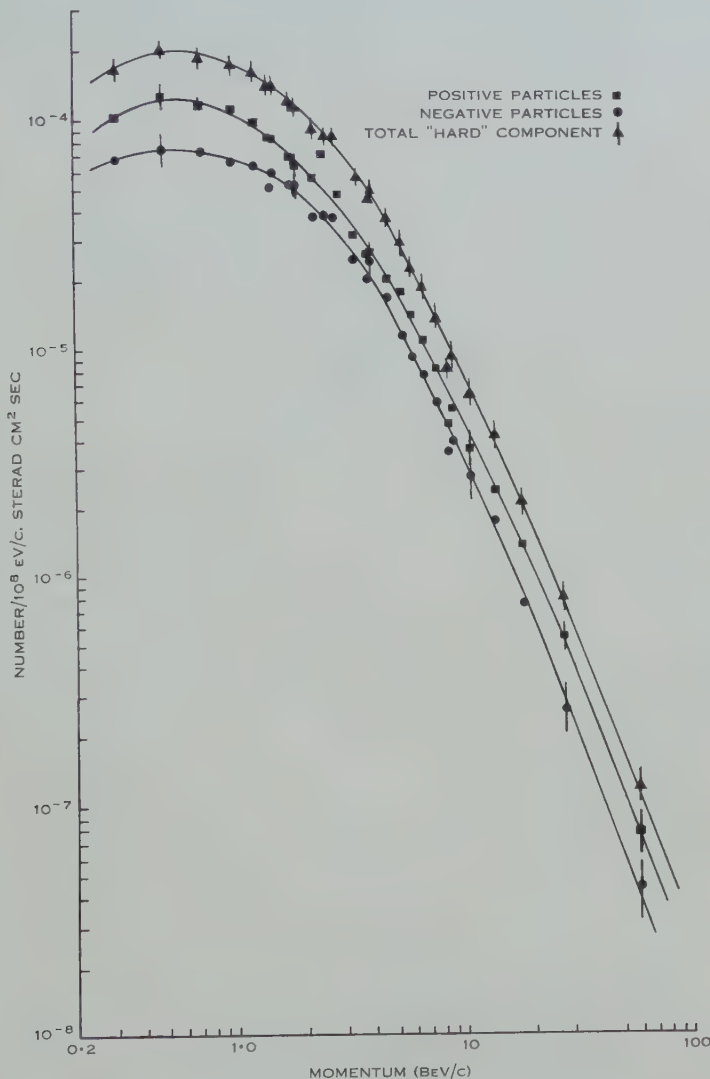


Fig. 1 (c).— 30° W. differential momentum spectra.

and 0.03 at 60° (Burbury and Fenton 1952), the accuracy of the absolute rate determination was insufficient for this to be detected.

A series of calculations has been made to determine whether the dependence of the total spectra on zenith angle can be explained solely in terms of the differing decay and momentum loss suffered by the μ -mesons in their paths through the atmosphere. The primary radiation, from which these particles

arise, is considered to be essentially isotropic at the top of the atmosphere since the geomagnetic latitude, 47°S , is at the knee of the latitude effect. In this section no account is taken of charge-sensitive phenomena within the atmosphere so that the μ -mesons are assumed to follow straight paths to sea-level.

The calculation falls into two parts; the tabulation of the survival probabilities and the derivation of a production spectrum which gives a reasonable and consistent description of the observed sea-level spectra.

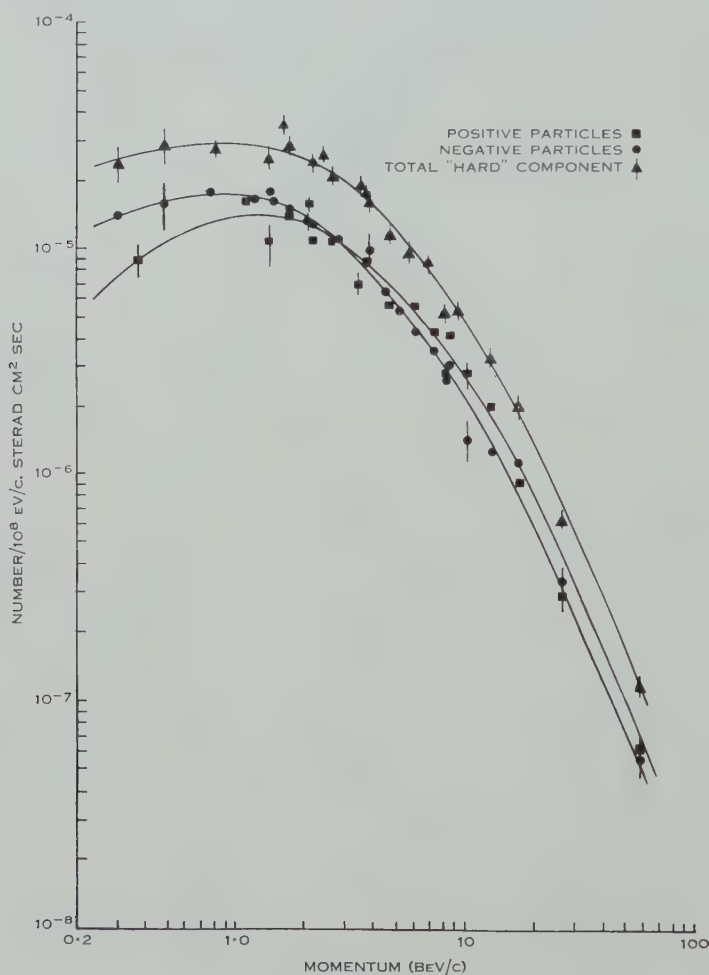


Fig. 1 (d).— 60°E . differential momentum spectra.

(ii) *Survival Probabilities*.—Denote by $w(x_0, p_0, \theta)$ the probability that a μ -meson will survive to sea-level and be recorded with momentum p_0 , at a zenith angle θ after production at a distance x_0 cm from sea-level, measured along its path. Then

$$w(x_0, p_0, \theta) = \exp \left[-\frac{1}{\tau c} \int_0^{x_0} \frac{dx}{p(x)} \right],$$

with $p(x)$ representing the momentum in units of μc at the distance x cm along the path. The function $p(x)$ may be evaluated by means of the range-momentum relation and a knowledge of the variation of absorber thickness along the track. A graphical representation of the distance-range relation, $x=f_\theta(R)$, was derived from pressure-altitude data provided by the Melbourne Meteorological Bureau. Here R g cm $^{-2}$ is the absorber thickness traversed to sea-level from x . The

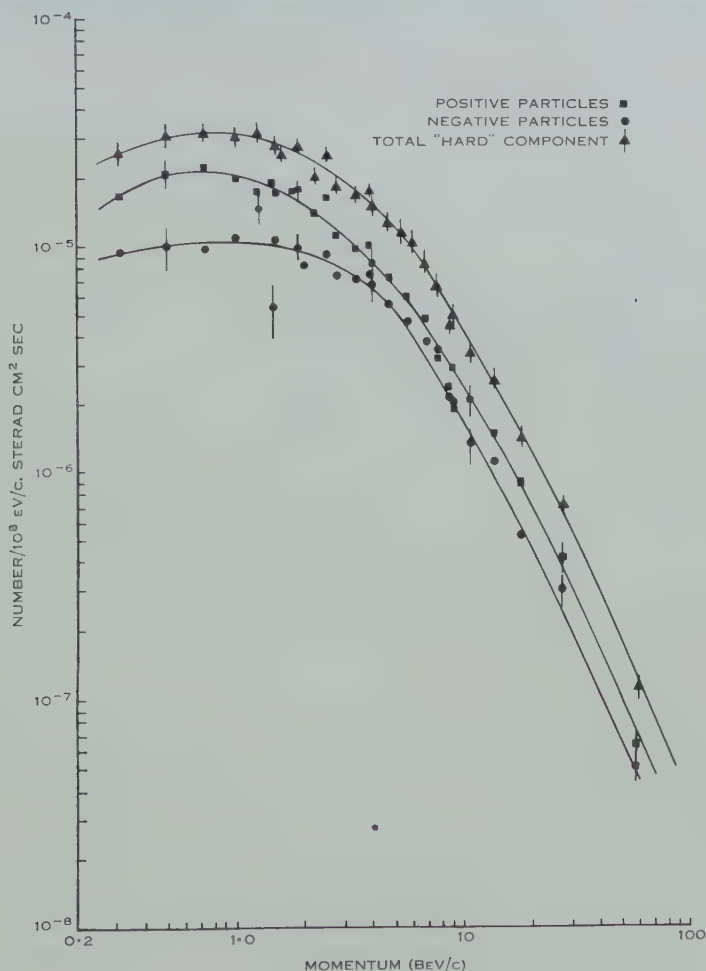


Fig. 1 (e).—60°W. differential momentum spectra.

numerical values of the function $p(f_\theta(R))$ used in the integration were obtained from the tabulated results of a range-momentum calculation (Fenton 1952).

The proper mean life τ and rest mass μ of the μ -meson were taken as 2.15×10^{-6} sec and 210 electron masses respectively, and the integral evaluated numerically at the zenith angles 0°, 30°, and 60°. A family of curves was obtained for each angle describing the variation of survival probability with height of production in g cm $^{-2}$, at eight values of the sea-level momentum p_0 .

(iii) *Single Layer Production*.—A reasonable description of the vertical spectrum has been found to be given by a process involving production at a single level near the top of the atmosphere (Euler and Heisenberg 1938 ; Janossy and Wilson 1946). When the decay of the mesons and the loss of momentum by ionization were taken into account, the production spectrum was shown to have the form of an inverse power law. Following the method of Janossy and Wilson (1946), a calculation was performed which indicated that the three sea-level spectra cannot be described simultaneously by a production process of

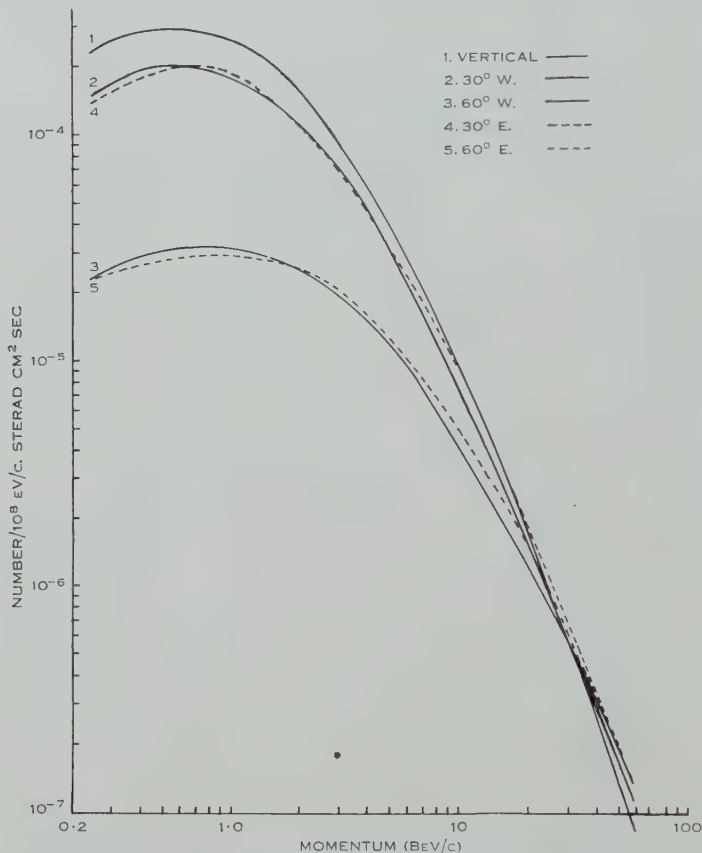


Fig. 2.—Zenith angle variation of the differential momentum spectrum—total hard component.

this nature. Moreover, these calculations showed, in agreement with conclusions reached by Sands (1950), that the mean height of production of μ -mesons reaching sea-level with a particular momentum decreases with decreasing momentum. For an adequate description, some form of extended production must be considered.

(iv) *Extended Production*.—The simplest model of an extended production process has been proposed by Sands (1950) who assumes that the dependences of production on depth and on momentum are separable. It is also assumed

that the production of mesons as a function of depth follows the same law as the absorption of the primary component. This is given by $\exp(-r/125)$ where r g cm⁻² is the depth from the top of the atmosphere measured along the path. As pointed out by Sands (1950) there is some justification for this separation at small values of r , but at greater atmospheric depths, where a secondary meson producing component may be appreciable, this assumption would break down.

It is known that the momentum dependent part of the production function is given to a good approximation by an inverse power law for momenta above several BeV/c. The behaviour of the function at low momenta may be obtained by fitting the results. The production spectrum obtained in this manner has the form

$$\begin{aligned} 0.147(p)^{-3.0} \cdot \exp(-r/125) & \quad \text{for } p > 17.6 \mu\text{c}, \\ 0.147(17.6)^{-3.0} \cdot \exp(-r/125) & \quad \text{for } p \leq 17.6 \mu\text{c}, \end{aligned}$$

where the intensity is in units of number of mesons (10⁸ eV/c)⁻¹ cm⁻² sec⁻¹. This spectrum differs from that found by Sands (1950). The spectrum derived by this author was based on observations of the low momentum intensity in the vertical direction at sea-level and at various altitudes. Although his spectrum is undoubtedly more reliable in the low momentum region it fails to give satisfactory agreement with the present results at high momentum.

The sea-level spectra were calculated by integrating the product of the production spectrum and survival probabilities over the atmosphere. These calculated spectra are compared with the measurements in Figure 3. Since only the total μ -meson component is considered it is felt that little error is made by combining the measurements in the eastern and western azimuths. The calculated spectra are normalized to the vertical observations at a momentum of 1.3 BeV/c.

Although the agreement is better than that achieved assuming single layer production it remains unsatisfactory. From Figure 3 it is seen that the vertical and 30° distributions, which were used in deriving the production spectrum, are adequately represented but that the calculated intensity at the 60° zenith angle is approximately 50 per cent. too low over a considerable region. At this angle, meson production as described by the depth dependent function, is already negligible at atmospheric depths such that mesons reaching sea-level are produced with momentum above the cut-off value, 17.6 μc .

A similar discrepancy was observed by Kraushaar (1949) who calculated the relative intensities of slow mesons at these zenith angles using a modified form of Sands' production spectrum. The above calculations have assumed that the mesons preserve the direction of the primary particles, however, it is known that at low momenta an appreciable spread is introduced due to both Coulomb scattering and the meson-producing events. No quantitative information is available on the angular distribution of the emitted mesons as a function of their momentum. However, taking the results of Brown *et al.* (1949) to refer to low momentum mesons, an approximate calculation was made to determine

the effect of this spread on mesons of 0.6 BeV/c at sea-level. The calculation shows that this effect is of the right order to account for the observed discrepancy. At high momentum these considerations are less important and the original assumptions will not be greatly in error.

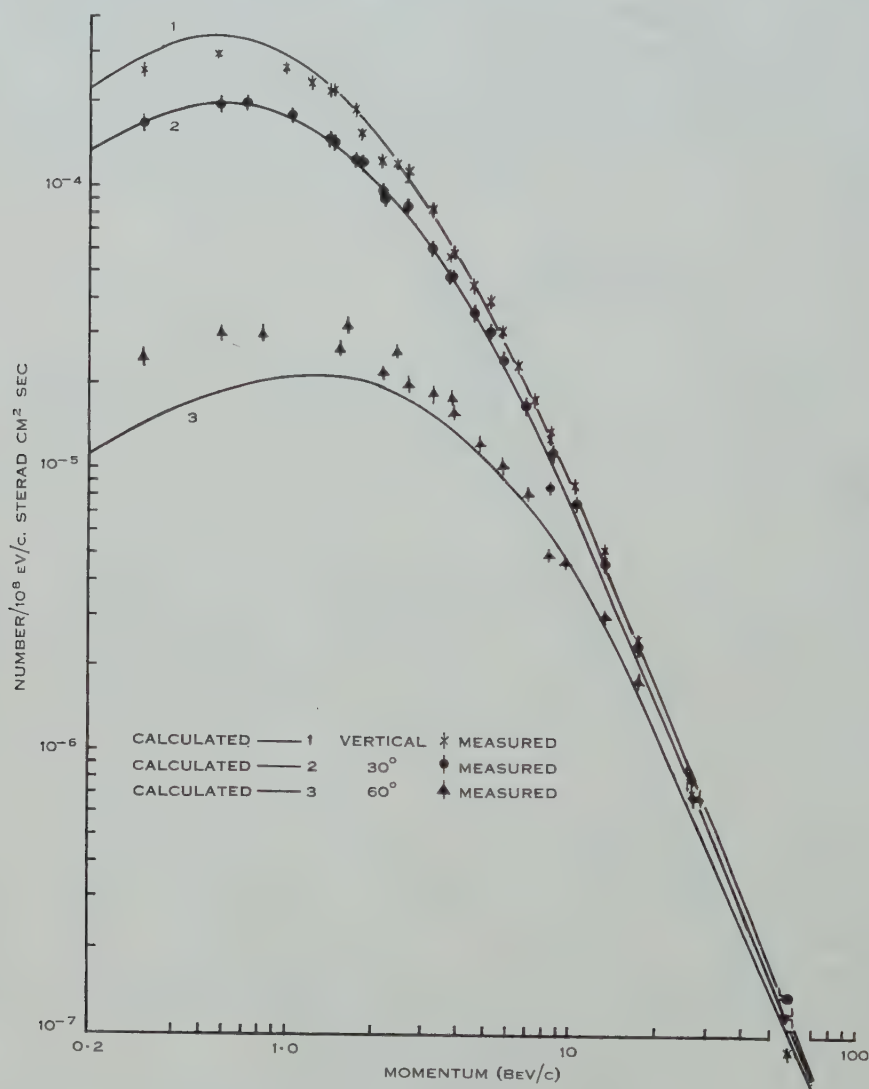


Fig. 3.—Differential momentum spectra.

(b) *The Positive and Negative Spectra and the Charge Ratio*

It may be seen from Figure 1 (a) that the positive and negative spectra obtained in the vertical direction have the same form and an approximately constant intensity ratio over the momentum range investigated. However, at inclinations to the vertical, Figures 1 (b)–(e), the positive and negative spectra differ in form and vary systematically as a function of zenith angle and

momentum. This effect manifests itself as large variations in the charge ratio (i.e. the ratio of the intensities of the positive and negative particles) and, since the results may be discussed more readily in terms of this quantity, we proceed to examine its behaviour.

The charge ratio in a given momentum interval may be obtained directly from the results. In order to determine the behaviour of this ratio as a function of momentum it is necessary to compromise between good momentum resolution and good statistics. This has been achieved by splitting the 1900 G measurements into three momentum intervals and the 13,500 G measurements into four intervals. The details of the momentum range and mean momentum (weighted by the distribution of particles within the range) are shown in Table 2. The groups are arranged in order of increasing momentum, L denoting a low field group and H a high field group. At the limiting momentum in each measurement there is an appreciable probability of contamination with particles of the opposite sign due to the finite resolution of the equipment. In order to avoid any uncertainty in the interpretation of the results, small deflexions in which the possibility of contamination is present are excluded from the analysis. Further, a small correction (less than 3 per cent. in the extreme case) is applied to the vertical measurements of the charge ratio to allow for the protons present in the penetrating component (Mylroi and Wilson 1951). No reliable data are available on this subject in the inclined directions but the subsequent error in the experimental charge ratios is expected to be less than 3 per cent.

The experimental values of the charge ratio are summarized in Table 2 and displayed in Figure 4. The results obtained in the vertical direction are shown together with the curve, representing the best estimate from available data, published by Owen and Wilson (1951). The measurements by these authors and by Beretta *et al.* (1953) possess considerably greater statistical accuracy than that attained in the present investigation; however, it will be shown later that our results agree satisfactorily with the work of these authors.

The main feature of interest in the present investigation is the behaviour of the charge ratio at inclined directions in the east-west plane. Examination of Figure 4 shows that large deviations from the values recorded in the vertical direction occur at each zenith angle. Several features are immediately evident. In the western azimuth the charge ratio progressively increases with decreasing momentum, the magnitude of this increase being accentuated at the greater zenith angles. The reverse effect occurs at easterly zenith angles and for sufficiently low momenta a negative excess is obtained.

Evidence of this phenomena has been obtained by several other authors (Groetzinger and McClure 1950; Beretta, Filosofo, and Sommacal 1952; Quercia and Rispoli 1953), all using magnetic lens techniques to separate the particles. Beretta and his co-workers made measurements of the positive excess of mesons with energies in the range $0.7\text{--}1.5$ BeV, in the vertical direction and at angles of 45° to the east and west. Their results show a positive excess of 0.41 ± 0.025 and 0 ± 0.023 for the penetrating component in the western and eastern azimuths. These results correspond to charge ratios of 1.52 ± 0.04 and 1.0 ± 0.03 respectively. These authors explained their results in terms of the curvature

TABLE 2
CHARGE RATIO OF THE PENETRATING COMPONENT AS A FUNCTION OF MOMENTUM

Group	Momentum Range (BeV/c)	Mean Momentum (BeV/c)	Charge Ratio								
			30°				60°				
			0°		W^+/W^-	E^+/E^-	W^+/E^-	E^+/W^-	W^+/W^-	E^+/E^-	W^+/E^-
L1	0.24-1.0	0.6	1.14 ± 0.07	1.61 ± 0.25	0.93 ± 0.15	1.19 ± 0.14	1.25 ± 0.15	1.90 ± 0.23	0.60 ± 0.11	1.37 ± 0.19	0.83 ± 0.15
L2	1.0-2.1	1.6	1.20 ± 0.07	1.40 ± 0.20	1.39 ± 0.22	1.43 ± 0.15	1.35 ± 0.16	1.65 ± 0.18	0.97 ± 0.13	1.08 ± 0.12	1.50 ± 0.19
H3	1.3-2.0	1.7	1.16 ± 0.10	1.44 ± 0.23	1.14 ± 0.13	1.27 ± 0.12	1.30 ± 0.13	2.20 ± 0.35	0.78 ± 0.13	1.26 ± 0.18	1.36 ± 0.24
H4	2.0-3.2	2.6	1.39 ± 0.12	1.39 ± 0.20	1.09 ± 0.12	1.30 ± 0.11	1.18 ± 0.11	1.70 ± 0.23	0.80 ± 0.11	1.07 ± 0.13	1.27 ± 0.19
L5	2.1-4.9	3.4	1.28 ± 0.08	1.25 ± 0.17	1.27 ± 0.19	1.23 ± 0.13	1.30 ± 0.13	1.51 ± 0.12	1.12 ± 0.12	1.34 ± 0.12	1.25 ± 0.12
H6	3.2-5.5	4.3	1.36 ± 0.12	1.19 ± 0.18	1.18 ± 0.13	1.24 ± 0.11	1.14 ± 0.11	1.34 ± 0.15	0.83 ± 0.10	1.14 ± 0.13	0.99 ± 0.13
H7	5.5-35	9.0	1.38 ± 0.12	1.50 ± 0.21	1.40 ± 0.13	1.31 ± 0.11	1.59 ± 0.14	1.29 ± 0.10	1.23 ± 0.10	1.17 ± 0.10	1.34 ± 0.11
Mean	1.266 ± 0.035	1.385 ± 0.054	1.202 ± 0.039	1.279 ± 0.045	1.300 ± 0.048	1.537 ± 0.061	0.979 ± 0.044	1.208 ± 0.049	1.244 ± 0.056

of the meson trajectories in the magnetic field of the Earth and concluded that the charge ratio at production is the same at these angles as in the vertical direction. The more recent experiments of Quercia and Rispoli (1953) were performed at a number of zenith angles, and their results show the general effect. Following the method of analysis proposed by Beretta, Filosofo, and Sommacal (1952), these authors could not explain the large residual fluctuations of the charge ratios and no definite conclusions could be reached.

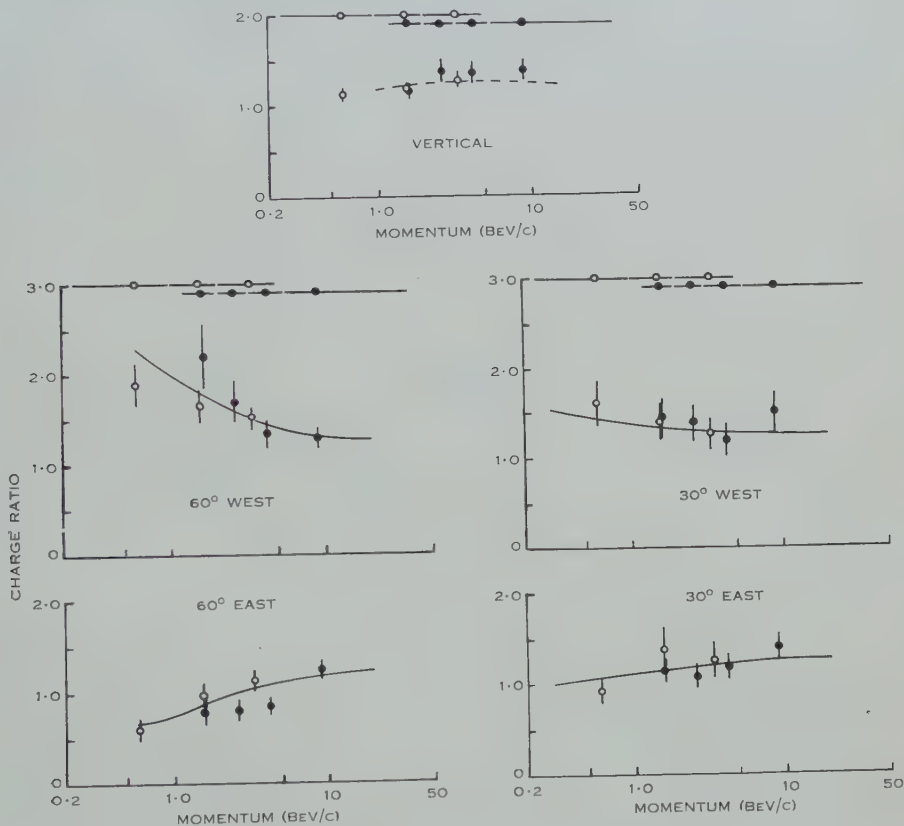


Fig. 4.—Charge ratio as a function of momentum.

○ 1900 G. ● 13,500 G.
 ----- From Owen and Wilson (1951).
 ——— Calculated.

Since the experiment was performed at a geomagnetic latitude of 47°S , it is extremely unlikely that an asymmetry in the primary radiation large enough to explain this effect could exist. However, due to the Earth's magnetic field, positive and negative mesons recorded with sea-level momentum p_0 at a zenith angle θ in the east-west plane will have trajectories of the form shown in Figure 5. It is assumed that, on the average, the primary component traverses the same atmospheric thickness (g cm^{-2}) in each case before producing mesons. Denoting by x_0 the distance measured along the trajectory from the point of

observation to the production level, the charge ratio of the mesons at sea-level may be written as

$$\frac{w(x_0^+, p_0, \theta)}{w(x_0^-, p_0, \theta)} \cdot \frac{K^+ p(x_0^+)^{-3}}{K^- p(x_0^-)^{-3}} = F(p_0, \theta, x_0^+, x_0^-) \frac{K^+}{K^-},$$

where $w(x_0, p_0, \theta)$ is the survival probability of mesons recorded at sea-level with momentum p_0 at the angle θ , $p(x_0)$ is the momentum at production, and $K^+(p)^{-3}$ and $K^-(p)^{-3}$ are the production spectra of the positive and negative mesons respectively. It is known from the vertical measurements that the charge ratio at production, K^+/K^- , is a slowly varying function of momentum. For the purpose of this discussion it may be considered constant.

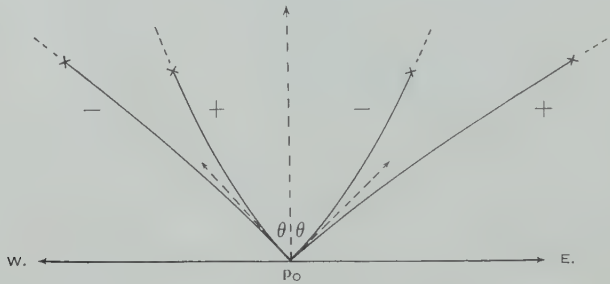


Fig. 5.—Meson trajectories in the Earth's magnetic field.

Reference to Figure 5 shows that in the eastern azimuth $x_0^+ > x_0^-$. Consequently

$$p(x_0^+) > p(x_0^-)$$

and

$$w(x_0^+, p_0, \theta) < w(x_0^-, p_0, \theta).$$

Therefore $F < 1$ and decreases as $x_0^+ - x_0^-$ increases.

Since the radius of curvature of the trajectory at any point is inversely proportional to the momentum at that point, it follows that F decreases as p_0 is decreased or θ increased. Similar considerations apply in the western azimuth. Here $x_0^+ < x_0^-$ so that $F > 1$ and increases as p_0 is decreased or θ increased.

The preceding discussion explains qualitatively the behaviour of the charge ratio in inclined directions. The vertical measurements are unaffected by the curvature since the trajectories of the positive and negative mesons are symmetrical about the zenith giving $x_0^+ = x_0^-$ and hence $F = 1$. It was pointed out by Beretta, Filosofo, and Sommacal (1952) that, in fact, all the meson trajectories are symmetrical about the zenith. Thus, for the charge ratios W^+/E^- and E^+/W^- the factor $F = 1$ under all circumstances and the observed values of these ratios at sea-level give a measure of the mean charge ratio at production.

The ratios W^+/E^- and E^+/W^- are collected in Table 2. The mean values shown represent suitably weighted averages over the spectrum. It is seen that those for W^+/E^- and E^+/W^- do not differ significantly from the mean 1.266 ± 0.035 recorded in the vertical direction. Examination of the individual

points indicates that, although large fluctuations about the mean value occur, there is no systematic variation with either zenith angle or momentum. A χ^2 test has been performed to ascertain whether these deviations were consistent with that expected from random fluctuation of the points. The following measurements have been tested against the curve published by Owen and Wilson (1951): the vertical ratios, the 30° W^+/E^- and E^+/W^- , the 60° W^+/E^- and E^+/W^- . The confidence levels obtained were 70, 90, 20, 50, and 8 per cent. respectively. All of these lie within the normal limits 5–95 per cent. Corresponding tests on the 30° W^-/W^- and E^+/E^- , and the 60° W^+/W^- and E^+/E^- gave the values 2, 2, 0.01, and 0.01 per cent., thus verifying the systematic deviation illustrated in Figure 4.

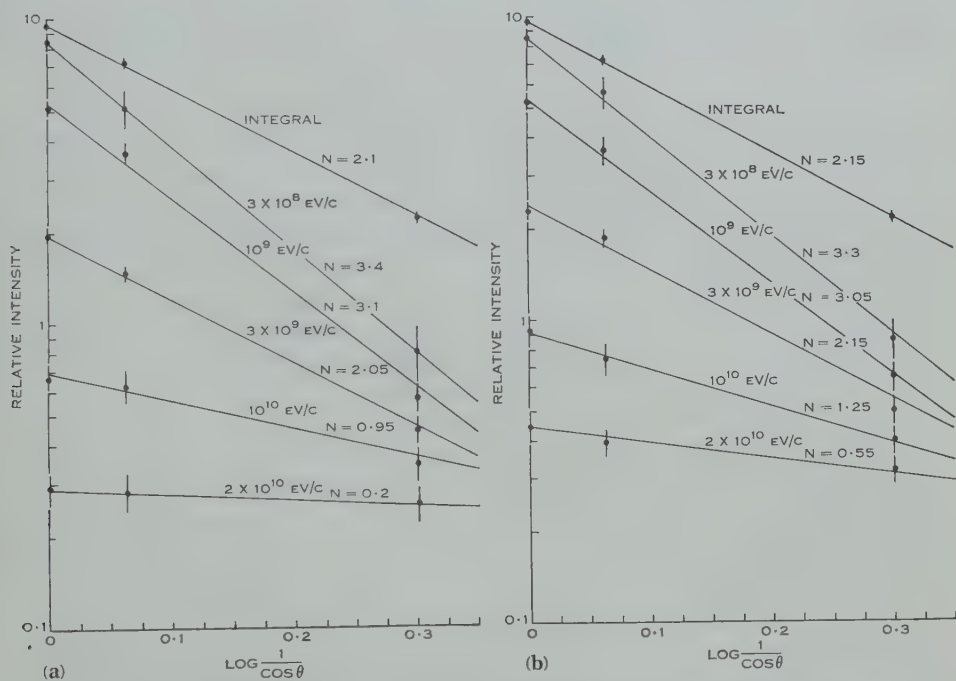


Fig. 6.—Variation of intensity I_θ with zenith angle θ at particular momenta, (a) to the east; (b) to the west.

$$I_\theta = I_0 \cos^n \theta.$$

As a final test, an approximate calculation of the charge ratios W^+/W^- and E^+/E^- was made by the method outlined earlier in the discussion. An approximate numerical evaluation of the path lengths x_0^+ and x_0^- allowed values of the factor F to be calculated. Assuming a value 1.25 for the charge ratio at production the curves shown in Figure 4 with the 30° and 60° measurements were obtained. The reasonable agreement with this approximate calculation, together with the previous tests, allows the conclusion that the behaviour of the charge ratio at inclined directions is explained as a secondary effect due to curvature of the meson trajectories in the magnetic field of the Earth.

(c) *The Intensity as a Function of Zenith Angle*

Previous work on this subject (Greisen 1942 ; Kraushaar 1949 ; Zar 1951 ; and others) has shown that in most instances the results could be fitted satisfactorily by a law of the form

$$I_{\theta} = I_0 \cos^n \theta,$$

where I_{θ} is the intensity at a zenith angle θ . A value of $n=2.1$ for the total penetrating component was obtained by Greisen (1942). However, delayed coincidence and anti-coincidence experiments on low momentum mesons (Kraushaar 1949 ; Zar 1951) have shown that the value of n may be as high as 3.3 in this region. No comprehensive measurements on the behaviour of n as a function of momentum are available. Although the statistical accuracy of the present measurements is not sufficient to provide a rigorous test of this expression, by assuming its validity the behaviour of the exponent n with momentum may be obtained.

The intensity of the total penetrating component and the differential intensity at five values of the momentum are plotted against θ in Figure 6. Figure 6 (a) refers to measurements made in the eastern azimuth and Figure 6 (b) to the corresponding measurements in the western azimuth. Values of $n=2.1$ and 2.15 are obtained for the total penetrating component, in good agreement with previous investigations. However, for mesons of 0.3 BeV/c the values $n=3.3$ and 3.4 are recorded (compare Kraushaar (1949) and Zar (1951)) and as the momentum increases the value of n decreases until the sea-level radiation at these latitudes becomes substantially isotropic above 20 BeV/c.

V. ACKNOWLEDGMENTS

The authors wish to thank Professor L. H. Martin for his interest in this work. They are indebted to Dr. H. D. Rathgeber, who suggested the investigation, for helpful discussions and advice on technical aspects throughout the project. The assistance of Mr. J. L. Rouse in the early part of the experiment is greatly appreciated.

VI. REFERENCES

- BERETTA, E., FILOSOFO, I., and SOMMACAL, B. (1952).—*Nuovo Cim.* **9** : 317.
 BERETTA, E., FILOSOFO, I., SOMMACAL, B., and PUPPI, G. (1953).—*Nuovo Cim.* **10** : 1354.
 BROWN, R. H., ET AL. (1949).—*Phil. Mag.* **40** : 862.
 BURBURY, D. W. P., and FENTON, K. B. (1952).—*Aust. J. Sci. Res.* **A 5** : 47.
 CARO, D. E., PARRY, J. K., and RATHGEBER, H. D. (1951).—*Aust. J. Sci. Res.* **A 4** : 16.
 EULER, H., and HEISENBERG, W. (1938).—*Ergebn. exakt. Naturw.* **17** : 1.
 FENTON, K. B. (1952).—Ph.D. Thesis, University of Tasmania.
 GREISEN, K. (1942).—*Phys. Rev.* **61** : 212.
 GROETZINGER, G., and MCCLURE, G. W. (1950).—*Phys. Rev.* **77** : 777.
 JANOSSY, J., and WILSON, J. G. (1946).—*Nature* **158** : 450.
 KRAUSHAAR, W. L. (1949).—*Phys. Rev.* **76** : 1045.
 MYLROI, M. G., and WILSON, J. G. (1951).—*Proc. Phys. Soc. Lond.* **A 64** : 404.
 OWEN, B. G., and WILSON, J. G. (1951).—*Proc. Phys. Soc. Lond.* **A 64** : 417.
 QUERCIA, I. F., and RISPOLI, B. (1953).—*Nuovo Cim.* **10** : 357.
 SANDS, M. (1950).—*Phys. Rev.* **77** : 180.
 WILSON, J. G. (1946).—*Nature* **158** : 414.
 ZAR, J. L. (1951).—*Phys. Rev.* **83** : 761.

HARMONICS IN THE SPECTRA OF SOLAR RADIO DISTURBANCES

By J. P. WILD,* J. D. MURRAY,* and W. C. ROWE*

[*Manuscript received April 2, 1954*]

Summary

The paper describes observations with a 40–240 Mc/s radio spectroscope leading to the discovery and preliminary investigation of harmonics in the spectra of sporadic bursts from the Sun. It is found that spectral features of bursts are commonly duplicated with a 2:1 frequency separation. The ratio is sometimes appreciably lower though apparently never greater. This and other features of the spectra are shown to support the hypothesis that the fundamental frequency corresponds to the natural plasma frequency of the corona in the vicinity of the source. By applying this result to a standard model of the corona, information on the position, velocity, and size of the sources is deduced. The results suggest that the generation of bursts may be associated with longitudinal plasma oscillations excited by fast streams of charged particles.

I. INTRODUCTION

Most of the energy that reaches us from the Sun at wavelengths in the radio spectrum between 1 and 20 m does so in short-lived sporadic bursts of seconds' or minutes' duration and in occasional storms which may continue for days. Since its discovery by Dr. J. S. Hey in 1942, the intense variable radiation has been studied extensively on a world-wide scale. But its origin is still largely unknown. We know from the application of the theory of ionized media to optical data that waves of such length must originate entirely in the outer layers of the solar atmosphere. We know from radio observations that the sources of some of the greater bursts move outwards through several hundred thousand kilometres of the solar corona during the few minutes in which they are received. We know also that the radio disturbances have a general association with visible activity, particularly large sunspots and solar flares. The experimental work which led to the present paper was undertaken with the object of looking for further clues on the physical nature of the radio disturbances.

The experiment consists of observing the spectrum of the high intensity radiation, as a function of time, over a wide continuous range of frequency. The range extends from 40 to 240 Mc/s, that is to say between wavelengths of 1.25 and 7.5 m. Some previous spectroscopic observations were made in the range 70–130 Mc/s (Wild 1950*a*, 1950*b*, 1951; Wild and McCreedy 1950); experience gained in this work emphasized the importance of extending the frequency range and helped to shape the design of the present experimental programme.

* Division of Radiophysics, C.S.I.R.O., University Grounds, Sydney.

On November 21, 1952, some four months after the start of regular observations, a large outburst of radio noise was recorded which revealed a new phenomenon. The record showed that spectral features in the lower part of the observed range were unmistakably duplicated at about double the frequency. In a preliminary report of this occurrence (Wild, Murray, and Rowe 1953), the authors pointed out that the duplication was almost certainly due to the emission of fundamental and second harmonic frequencies from a common source. In view of subsequent evidence this conclusion now seems definite. Indeed it appears that harmonics are received in a considerable proportion of all sporadic bursts.

In the present paper we give details of observations on the harmonic effect (Section III), suggest an interpretation of the results (Section IV), and show how the data may be used to study the location, speed, and size of the sources responsible for generating sporadic bursts (Section V). The conclusions of the paper are summarized in Section VI. Before considering these topics a brief outline is given of the experimental method.

II. THE EXPERIMENTAL METHOD

(a) *Choice of Site*

The choice of site for this investigation was influenced mainly by the requirement of avoiding interfering signals, especially those from high frequency radio transmitters. The observing station was set up near Dapto, N.S.W., some 50 miles to the south of Sydney. It is effectively screened from Sydney by nearby mountains.

(b) *The 40-240 Mc/s Spectroscope*

A detailed description of the instrument is beyond the scope of the present paper. Here we shall consider only the main features which determine its capabilities and the type of record produced.

The spectroscope is represented, in considerably simplified form, by the block diagram in Figure 1. Spectra are obtained by rapidly tuning through the range with a receiver of small bandwidth. The whole range is covered by three separate receiving units sweeping in succession through ranges of 40-75, 75-140, and 140-240 Mc/s. The three ranges have their outputs connected to common displays.

Each receiving unit consists of a broad-band rhombic aerial connected to a swept-frequency superheterodyne receiver. The receivers are tuned by continuously rotating tuning condensers connected to a common shaft which is driven by a motor at 2 r.p.s. Each range is swept in one-quarter of a revolution, the complete range in three-quarters. Thus a complete spectrum is swept out in $\frac{3}{8}$ sec and two complete spectra are obtained each second.

The receivers are controlled by an electronic sequence switch which allows the output of each receiver to be passed to the display only during its operative periods.

The frequency resolution of the instrument is determined by the response curve of the intermediate-frequency amplifier whose bandwidth between half-

power points is 0.5 Mc/s. The output time-constant, which must necessarily be short in view of the rapid sweep, is about 10^{-3} sec. The noise factor of the receivers varies slightly over the range with an average value of about 10.

The aerials are equatorially mounted and motor driven to follow the Sun. Their effective areas are between 5 and 10 m².

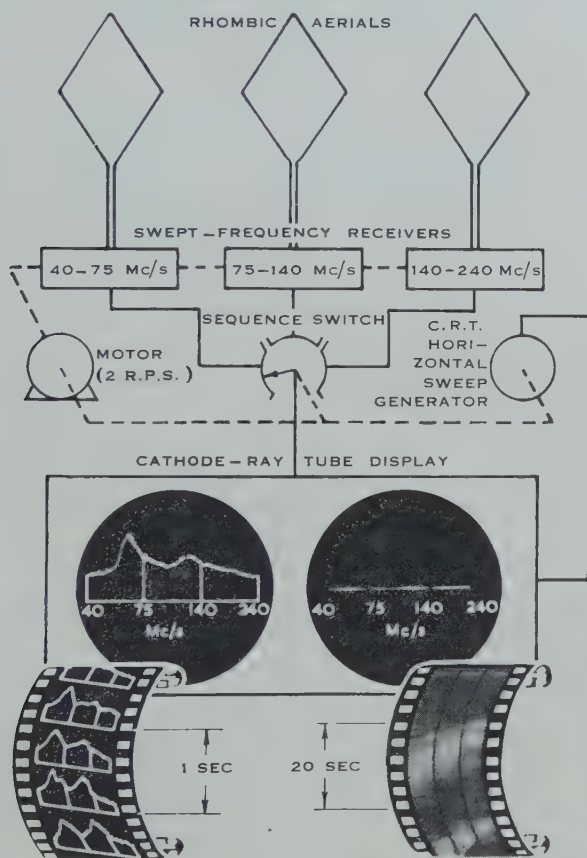


Fig. 1.—Simplified block diagram of the 40-240 Mc/s spectroscope.

(c) The Display and Recording Unit

Two types of recording display are used, each having a 12-in. cathode-ray tube and a camera operated with continuously moving 35-mm film.

One display is similar to that used for the previous observations (Wild and McCready 1950) and corresponds to the "A scan" of radar sets; frequency is displayed horizontally and receiver output vertically. This arrangement is suitable for precise measurements of absolute intensity but suffers from the disadvantages that (i) the rate of film consumption (36 in/min in our case) is excessively high for continuous operation and (ii) the process of reducing records to a useful form is too laborious for handling extensive data.

For continuous operation the second display is used in which frequency is again displayed horizontally but receiver output is registered by modulating the intensity of the spot. Since the response of photographic film is approximately logarithmic it is possible to record intensities of vastly different magnitudes. Using Ilford HP3 film it is found that aerial temperatures in a 1000 : 1 range can be satisfactorily accommodated. With a film consumption of 1 in/min (120 traces per inch), successive traces are recorded with no overlap.

(d) Calibration

Frequency calibration is provided by injecting into the receiver signals from a standard signal generator at frequency intervals of 5 or 10 Mc/s. The probable error associated with the frequency measurement of a sharply defined feature is determined mainly by the reading accuracy and is estimated to be $\pm \frac{1}{2}$ per cent.

Intensity calibration is provided by switching on temperature-limited noise diodes which are *permanently* connected across the aerial terminals of each receiver. The noise-source impedance is then provided by the aerial itself. This arrangement has two advantages: firstly, the anode-cathode capacity of the diode (which would be difficult to tune out over a broad frequency band with the conventional method of substitution) forms part of the input tuned circuit, thus eliminating calibration errors due to this cause; and secondly, if calibration is performed with the aerial directed towards the undisturbed Sun, the power injected by the noise generator corresponds to the excess power above cosmic noise, which is the quantity required to be measured. The standard calibration procedure consists of injecting power at seven levels each separated by a 2 : 1 interval. By this means the records are directly calibrated for levels of flux density between 5×10^{-21} and $3.2 \times 10^{-19} \text{ Wm}^{-2} (\text{c/s})^{-1}$; for higher levels it is necessary to extrapolate according to the receiver gain law.

Time marks are inserted on the record by photographing a clock beside the display which is illuminated by a flash at 1-min intervals.

The complete calibration procedure is performed at least once per day.

(e) The Reduction of Records

Spectra which vary with time are conveniently presented in the form of intensity-frequency-time diagrams in which intensity is shown by contours in the frequency-time plane. Such *dynamic spectra* were used to present the results of the previous observations. Experience then gained indicated that a practical limit to the amount of data that could be examined was set by the labour involved in producing the diagrams. However, by using the intensity-modulated display, we obtain the information in a form from which the salient features of the dynamic spectrum can be immediately recognized (see Plate 1 for examples). It is a simple process to trace these features and redraw them on a proper scale of frequency. These "sketches" are adequate for many purposes.

When higher precision is required (e.g. for the detailed study of the harmonic effect) it is more satisfactory to use an A-scan record and deduce the fully

calibrated profiles and dynamic spectrum. When only the intensity modulated record is available, it is necessary to analyse the film with a high resolution microphotometer.

III. THE OBSERVATIONS

Solar observations were started in August 1952 over the frequency range 40–140 Mc/s, and in November 1952 over the complete range 40–240 Mc/s. The observations described here cover the period August 1952 to August 1953. The equipment was operated for about 1000 hr during the year, preference being given to periods of solar activity.

In addition to several days of continuous activity ("noise storms") we recorded at other times several hundred bursts of sporadic occurrence. In the present paper we shall be concerned only with the sporadic bursts. In describing their spectra the following terms, previously introduced by Wild and McCreedy (1950), will be adopted and where necessary extended:

Spectral Type II.—The type of dynamic spectrum exhibited by bursts or "outbursts" of some minutes' duration in which the spectral features drift slowly, though perhaps irregularly, in the direction of decreasing frequency. The typical drift rate is of the order of $\frac{1}{4}$ Mc/s per sec. These bursts often occur at the time of solar flares.

Spectral Type III.—The type exhibited by sporadic bursts of a few seconds' duration in which the frequency of maximum intensity drifts rapidly in the direction of decreasing frequency at a rate of about 20 Mc/s per sec.

In the current series of observations the great majority of sporadic bursts were found to belong to one of these two spectral classes; four type II bursts were recorded and several hundred type III, the latter occurring mainly in small groups or compact clusters lasting for about 1 min.

The coexistence of first and second harmonics has been recognized in both classes of spectra—in two of the four observed type II bursts, and in 20 type III bursts. The spectrum of many other type III bursts suggested that harmonics might have been present though their certain recognition was masked by the large bandwidth of the bursts.

(a) *Harmonics in Type II Spectra*

The two recorded harmonic outbursts of spectral type II occurred on November 21, 1952 and May 5, 1953. The earlier one is the larger and more complete disturbance but the later one has yielded more exact data because an A-scan record was obtained. Details of the two outbursts are as follows.

(i) *The Outburst of November 21, 1952*

General data.—The outburst lasted from 23 hr 50 min to 24 hr 05 min U.T. and was followed immediately by a noise storm lasting for about 2 hr. The disturbance was accompanied by a large solar flare starting at 23 hr 45 min located above a big spot group near the centre of the Sun's disk. It was also accompanied by a radio fadeout and geomagnetic crochet.

The dynamic spectrum.—The outburst was recorded on the intensity-modulated display and the record reduced photometrically. Part of the record is shown in Plate 1 (a) and the dynamic spectrum in Figure 2 (a). It consists of

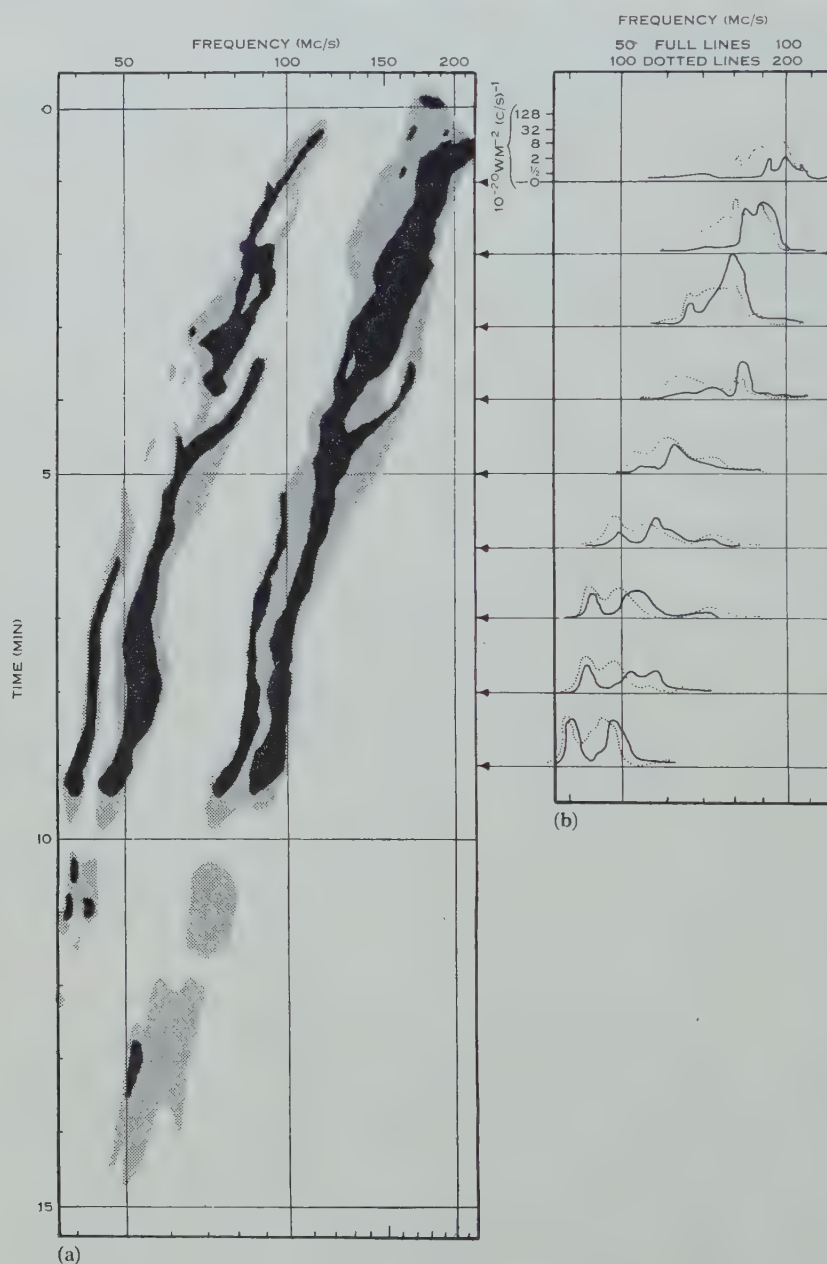


Fig. 2.—The outburst of November 21, 1952, 23 hr 50 min U.T. (a) The dynamic spectrum. The intensity contours correspond to levels of approximately 5 and 20 $\text{Wm}^{-2} (\text{c/s})^{-1}$. (b) Profiles at 1-min intervals. The second harmonic is shown dotted and displaced in frequency by a factor of 2.

two main bands, widely spaced in frequency, which drift at a rate typical of type II spectra. Each band shows a fine structure which for much of the time consists of two well-defined peaks. Two features are of special interest. Firstly, the structure of the two bands is strikingly similar; peaks appearing in one band seem to be duplicated simultaneously in the other. Secondly, although the frequency of each band drifts over a range of at least 3 : 1, the frequency ratio of corresponding features in the two bands remains approximately constant and lies within a few per cent. of 2. The first of these features was interpreted as signifying that the two main bands were emitted from the same source; the second that the duplicity was due to radiation from the source at a fundamental frequency and its second harmonic.

The profiles.—Figure 2 (b) shows the instantaneous spectral profiles of the two main bands taken at 1-min intervals; the bands are plotted in superposition such that a frequency f in the fundamental (full line) coincides with a frequency $2f$ in the harmonic (dotted line). When compared in this way the two sets of profiles are by no means identical in shape. One feature seems to be systematic: peaks in the harmonic band always lie slightly to the left of corresponding ones in the fundamental, i.e. the frequency ratio is consistently less than 2. The measured ratios lie between 1.96 and 1.99 for sharp peaks, and as low as 1.90 for the more diffuse peaks. Peak amplitudes in the two bands are of comparable magnitude.

(ii) *The Outburst of May 5, 1953*

General data.—The type II burst lasted from 04 hr 59 min to 05 hr 02 min U.T. It was preceded by a cluster of type III bursts and other activity starting at 04 hr 55 min, and was followed by a general increase in level at the lower frequencies lasting from 05 hr 04 min to 05 hr 08 min. No flare observations are available but the outburst accompanied a partial radio fadeout. Visible activity on the disk was confined almost entirely to western heliographic longitudes, between 30° and the limb.

The dynamic spectrum.—The outburst was recorded on both displays, the A-scan record being used exclusively for deriving the data presented below. The dynamic spectrum is shown in Figure 3 (a). Both the duration and rate of frequency drift are smaller than for the previous outburst. The spectrum is simpler and, except near the finish, each harmonic band consists of a single peak.

The profiles (Fig. 3 (b)).—Owing to the greater precision of the A-scan record, the profiles have been plotted on a larger scale and at more frequent intervals (15 sec). The profiles of the fundamental peaks are seen to be markedly asymmetrical, the low frequency edge being the steeper of the two. Indeed in the last four profiles, the slope of this low frequency "cut-off" is indistinguishable from that of the receiver's response curve, indicating that the slope on the true spectrum was too steep to be resolved by the instrument. The harmonic band is more symmetrical but on the average the low frequency edge is again the steeper. With the passage of time the ratio of peak frequencies gradually

increases from about 1.7 near the start to 2.00 ± 0.01 at the finish and the amplitude of the fundamental relative to the harmonic increases from about 0.5 to 10 .

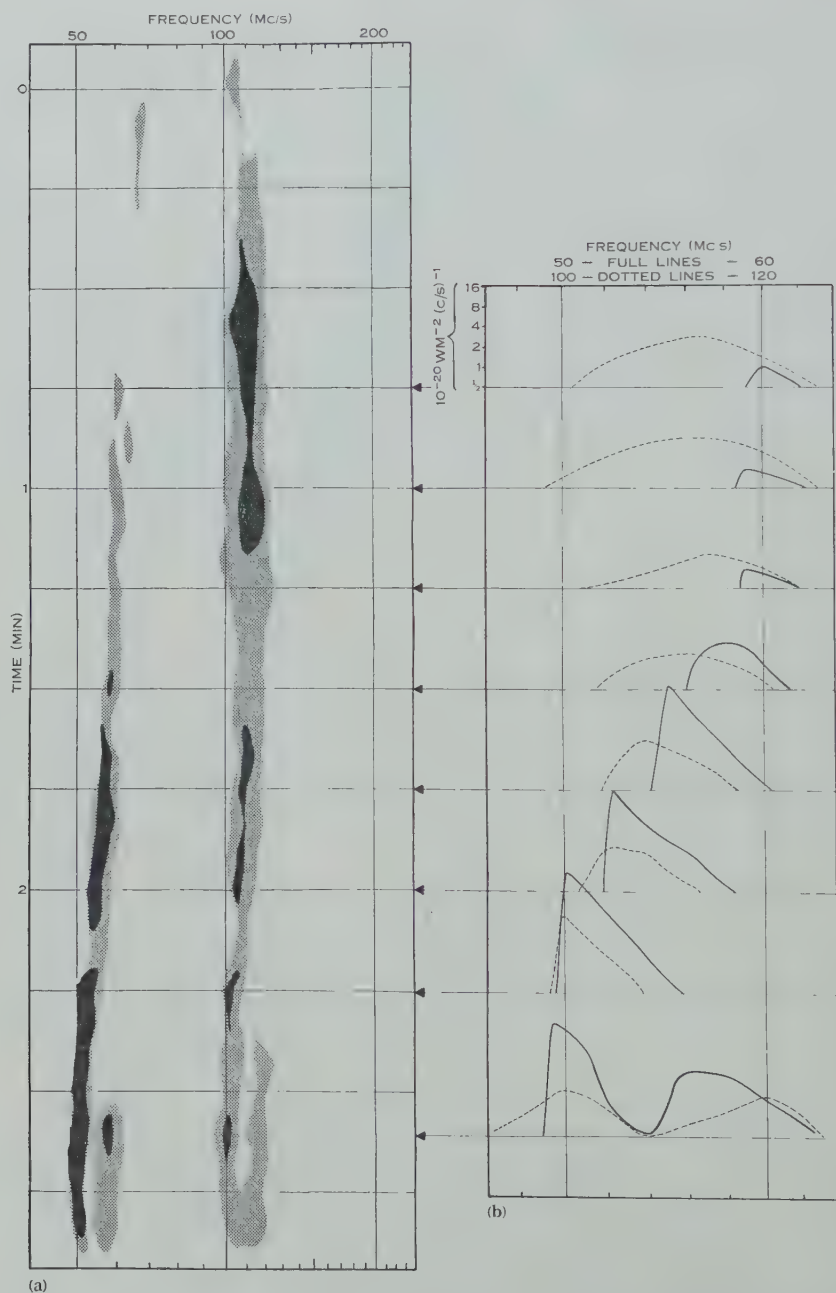


Fig. 3.—The outburst of May 5, 1953, 04 hr 59 min U.T. (a) The dynamic spectrum (contour levels as in Fig. 2), (b) profiles at 15-sec intervals. Note that scales are different from those in Figure 2.

(b) Harmonics in Type III Spectra

The fast-drift, short-lived bursts of spectral type III, first observed with the previous equipment (Wild 1950*b*), have been found to constitute the great majority of all bursts observed at times other than noise storms. With the wider frequency range of observation it is now possible to give what seems to be a complete description of their spectral characteristics. The new finding of

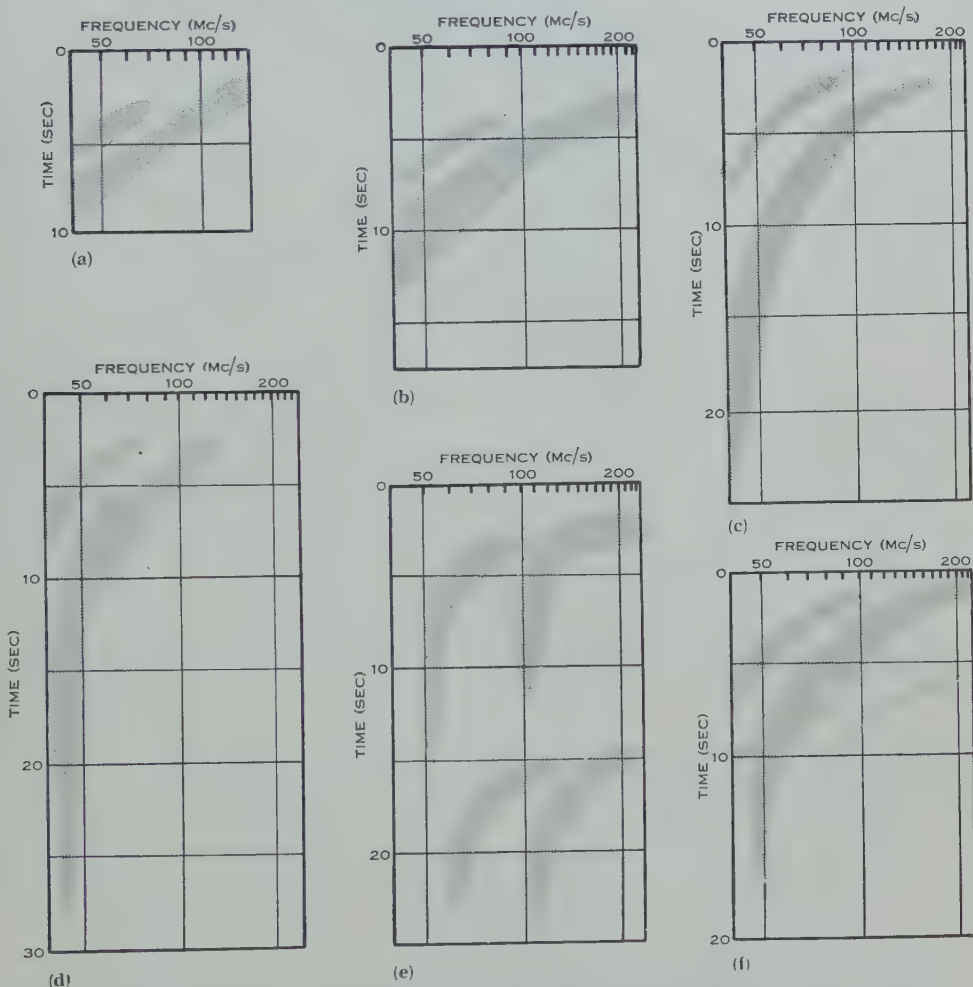


Fig. 4.—Dynamic spectra of harmonic type III bursts. Times (U.T.): (a) October 3, 1952, 23 hr 21 min; (b) June 7, 1953, 01 hr 08 min; (c) June 5, 1953, 01 hr 35 min; (d) June 5, 1953, 01 hr 37 min; (e) January 14, 1953, 06 hr 07 min; (f) June 5, 1953, 01 hr 32 min.

immediate interest is the recognition of dynamic spectra consisting of two similar formations in which the features of one are duplicated at about double the frequency. Following the evidence given above for type II spectra, there seems little doubt that this effect is again due to the emission of a fundamental frequency and its second harmonic.

Figure 4 shows examples of harmonics in type III bursts. In cases (a), (b), and (c) both fundamental and harmonic bands are seen to drift off the low frequency edge of the observed range, the rate of frequency drift decreasing continuously with time. In cases (d), (e), and (f), however, the frequency drift decreases so rapidly at the lower frequencies that one or both bands become

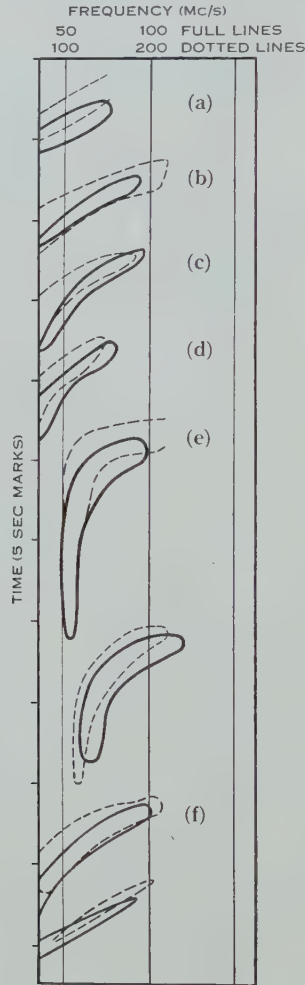


Fig. 5.—Comparison of the fundamental and second harmonic bands of the type III bursts shown in Figure 4.

almost stopped at a fixed frequency: the spectrum finishes in a vertical "tail". In all cases the fundamental and harmonic parts of the dynamic spectrum are similar in shape although the former tends to be of narrower relative bandwidth ($\Delta f/f$).

In the type II spectra considered above it was shown that the ratio of peak frequencies in the two bands never exceeded 2 but was often slightly lower. A

similar effect is demonstrated for type III spectra in Figure 5 in which relevant portions of the spectra in Figure 4 have been replotted with the harmonic band displaced two-to-one in frequency. There is seen to be a consistent displacement of the harmonic band to the left of the fundamental, again indicating a frequency ratio less than 2. In the great majority of cases the ratio lies between 1.85 and 2.00.

Of the several hundred type III bursts observed during the year, the recognition of harmonics was considered certain on 20 occasions. In addition there were many doubtful cases. Certain recognition is often made difficult owing to the combination of wide bandwidth and rapid frequency drift, which may result in the merging of the two bands. In cases where the bursts show tails, however, it is possible to decide without ambiguity whether or not the two bands are present. Of 19 such cases recorded, 12 were found to have the two bands. If it is assumed that the cause of harmonic production is independent of the cause of tail production, this result suggests that the proportion of bursts in which fundamental and harmonic are both of detectable magnitude may exceed 50 per cent.

(c) Summary of Observational Results

The main observational results are summarized below.

(1) The dynamic spectrum of sporadic bursts, both type II (slow drift) and type III (fast drift) sometimes show simultaneous duplication of features separated by a frequency ratio of about 2 : 1. The duplication is attributed to the emission of both fundamental and second harmonic frequencies from a common source.

(2) Bursts showing detectable fundamental and harmonic bands in their spectra account for a considerable proportion of all bursts, perhaps 50 per cent. or more. At the higher frequencies, in the vicinity of 200 Mc/s, it is not unreasonable to surmise that the greater part of radiation received in sporadic bursts is due to second-harmonic emission.

(3) The intensity of the second harmonic may be comparable with, or even greater than that of the fundamental. No third and fourth harmonics have been detected; had they been present with an intensity one-tenth of that of the second harmonic, they would certainly have been detected on several records.

(4) There appears to be a general rule that the ratio of peak frequencies never exceeds 2 but is often slightly lower.

(5) Detailed analysis of one outburst of type II for which sufficiently precise data were available indicated that the low frequency edge of the fundamental band may be extremely sharp and that the relative amplitude of the fundamental increases as the frequency ratio approaches 2.

IV. INTERPRETATION

(a) General Inferences

The first three conclusions listed above imply that the sporadic solar radiation is rich in second harmonics. This surely means that the emitting process is one involving oscillations of charge, as distinct from non-periodic accelerations such

as those of thermal motions in the absence of a magnetic field. To produce a harmonic the oscillators must be non-linear. Also, since an even harmonic is observed, the emitted waveform must be asymmetrical in the sense that values of the electric vector at two instants of time separated by half a cycle are not equal and opposite. This asymmetry may provide an important restriction on the types of processes admissible.

In some bursts, especially those of type II, it was noted that the width of each band is extremely narrow, perhaps only 2 or 3 per cent. of the mid frequency, between half-intensity points. We infer that in these cases all the oscillatory charges which constitute the source oscillate at roughly the same frequency. In other words there appears to be some proper frequency controlling the oscillations.

We know of two classes of proper frequency for oscillations in an ionized medium: the gyro frequencies of charged particles in a magnetic field, of which the electron gyro frequency is the most relevant in the present problem; and the plasma frequency.

(i) *The Electron Gyro Frequency*, $f_H = eH/2\pi mc$.—Here e and m denote the electronic charge (e.s.u.) and mass, H the magnetic field, and c the velocity of light. It has been shown by Schwinger (1949) and others that electrons gyrating in a magnetic field generate harmonics when the orbital velocity approaches the velocity of light. The difficulties associated with the escape from the solar atmosphere of radiation at the fundamental of the gyro frequency (Ryle 1948) are well known, though it is possible that the harmonics $2f_H$, $3f_H$, . . . could escape (Roberts 1952). Indeed Roberts suggested that spacings between harmonics offered a possible experimental test for the gyro theory of generation; he pointed out that adjacent harmonics would be in the ratios 3 : 2, 4 : 3, 5 : 4, etc., but not 2 : 1. The observation of harmonics at a 2 : 1 spacing now provides a negative answer to this test and indicates that a process is required in which fundamental frequencies can escape.

(ii) *The Plasma Frequency*, $f_0 = e\sqrt{N/\pi m}$.—Here N is the electron density of the medium. In the presence of a magnetic field, not considered here, the plasma frequency is split into three components one of which is f_0 .

Several authors (see for instance Shklovsky 1946; Martyn 1947; Bohm and Gross 1949; Jaeger and Westfold 1949) have suggested oscillations at the plasma frequency as the source of high intensity solar radio noise although no complete theory has yet been given. Difficulties are again encountered regarding escape because the plasma frequency coincides with the critical frequency at which the refractive index reduces to zero. If emission takes place from a localized region in which only the plasma frequency is excited, the radiation can escape only within an infinitesimally narrow cone normal to the surface of zero refractive index. However, there is no such escape restriction on the higher harmonics.

Let us now suppose that the excited region generates not merely the plasma frequency but rather a narrow band of frequencies about the plasma frequency together with their harmonics. Since propagation can take place at frequencies

above the critical frequency, the received spectrum would consist of all second and higher harmonic frequencies generated but only the high frequency part of the fundamental band. This is strongly suggestive of the observed harmonic spectra and seems to offer a natural explanation both of the reduction of the harmonic ratio below 2 and the sharp cut-off in the fundamental band. We now examine this interpretation in more detail by considering the spectrum from a point source in a model corona.

(b) *The Escape of Frequencies near the Plasma Frequency from a Model Corona*

The approach to be adopted may be summarized as follows. We consider a localized source in the solar corona situated in a region specified by an electron density N . We suppose that if a probe connected to a radio spectroscope were inserted into this region the resulting spectrum would consist of a narrow-band peak centred on the plasma frequency f_0 and others of similar shape but lower amplitude at the harmonic frequencies $2f_0, 3f_0 \dots$. We refer to this spectrum as the *natural spectrum*. Next we consider the propagation of energy through the solar atmosphere and calculate the fraction of the source brightness capable of reaching a terrestrial observer. The fraction transmitted is a function of frequency which we call the *propagation characteristic*. Finally we deduce the *received spectrum* by multiplying the natural spectrum by the propagation characteristic.

For the electron-density distribution in the corona we shall assume the idealized, spherically symmetrical model given by the Baumbach-Allen formula. The effects of coronal irregularities and magnetic fields will be neglected for the present calculation but will be considered in Section V. We shall be concerned merely with illustrating the type of result obtained and it will be sufficient to consider the case of a fixed point source whose height we choose to be at the 50 Mc/s plasma level (2.3×10^5 km above the photosphere in the Baumbach-Allen corona).

(i) *The Natural Spectrum*.—For present purposes it is convenient to consider a natural spectrum of definite form, though its choice is relatively unimportant. We shall assume the profile of each harmonic to be symmetrical, on a logarithmic frequency scale, about the plasma frequency and to have the shape of the Gaussian error curve. The assumed spectrum is shown in Figure 6 (a). The bandwidth has been chosen to be comparable with observed profiles and the amplitude of the second harmonic has been set arbitrarily to one-tenth that of the fundamental.

(ii) *The Propagation Characteristic*.—The propagation and escape of radio waves from point sources within the Baumbach-Allen corona has been treated in detail by Jaeger and Westfold (1950). The problem consists of calculating (1) the paths of rays between source and observer, in general these paths are curved owing to the steady increase with height of refractive index in the corona; and (2) the absorption along each path.

Figure 7 shows the paths of rays of different frequencies escaping from a source at the 50 Mc/s plasma level. The lowest frequency capable of escape is

the plasma frequency itself which can be propagated along the radial ray only. Higher frequencies can escape in directions contained within a sharply defined cone about the radial direction; the higher the frequency the wider the cone.* Thus to a terrestrial observer the propagation characteristic depends on the position of the source on the Sun's disk. Figure 6 (b) shows the derived characteristic for various source angles θ (defined in the inset below Figure 6 (a)). The effect of absorption is included and accounts merely for the slight drop in values of transmission at frequencies immediately above the cut-off frequency. Apart from this small effect the cut-off is perfectly sharp. As implied by Figure 7, the cut-off frequency coincides with the plasma frequency for a source at the centre of the disk and increases with the source angle.

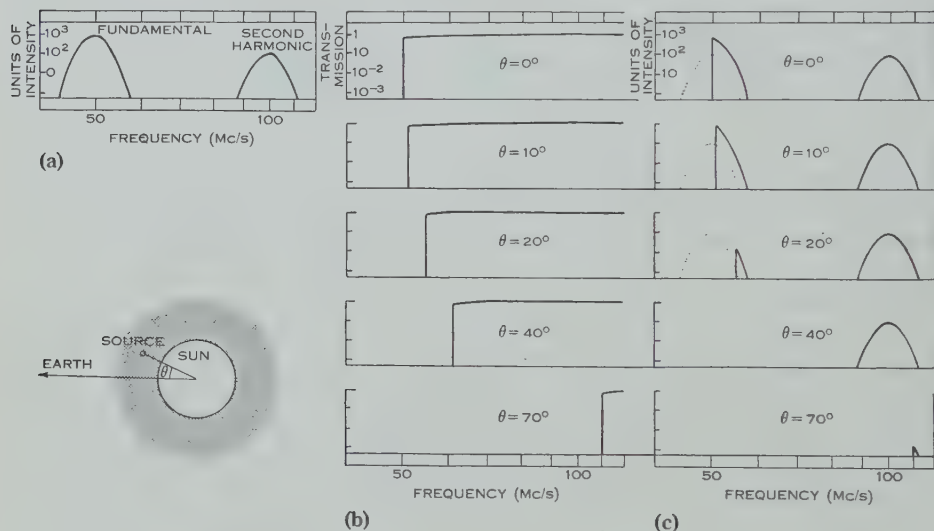


Fig. 6.—The derivation of spectra for an assumed model (see Section IV). (a) The assumed natural spectrum for a point source, located at the 50 Mc/s plasma level of the solar corona, emitting harmonics of the plasma frequency; (b) the calculated propagation characteristic (from Jaeger and Westfold) for radiation escaping from the 50 Mc/s level at various angles θ (see inset below (a)); (c) the received spectrum derived from (a) and (b).

(iii) *The Received Spectrum.*—Combining the data of Figures 6 (a) and 6 (b) we obtain the received spectra of Figure 6 (c). For a source at the centre of the disk ($\theta=0^\circ$), the ratio of peak frequencies of the fundamental and second harmonic is exactly 2, but the low frequency half of the fundamental band is completely cut off. For $\theta=10^\circ$ the ratio is reduced to 1.96 and for $\theta=20^\circ$ to 1.90. In the latter case the amplitude of the fundamental is greatly reduced, and at $\theta=40^\circ$ it is no longer present. For these values of θ the second harmonic

* Each ray path shown in Figure 7 is that which leaves the source tangentially to the spherical strata of the assumed atmosphere. In general the "outermost" ray (i.e. that which emerges from the atmosphere at the greatest angle) is one which leaves the source in a direction having a slight inward component. The difference in the angles of emergence of these two types of limiting ray is negligible except for very oblique emergence.

is unaffected, but for sources near the limb (e.g. $\theta = 70^\circ$) it too becomes modified by the cut-off characteristic.

(iv) *Comparison with Observations.*—The derived spectra of Figure 6 (c) illustrate how localized emission at harmonics of frequencies around the plasma frequency can explain the following characteristics observed in type II and type III spectra :

- (1) The normal occurrence of harmonic peak-frequency ratios slightly less than 2.
- (2) The sharp low frequency cut-off observed in the fundamental band.
- (3) The reduction in amplitude of the fundamental peak relative to the harmonic as the peak-frequency ratio diminishes.

There is seen to be a marked resemblance between the derived profiles and the detailed observed profiles shown in Figure 3.

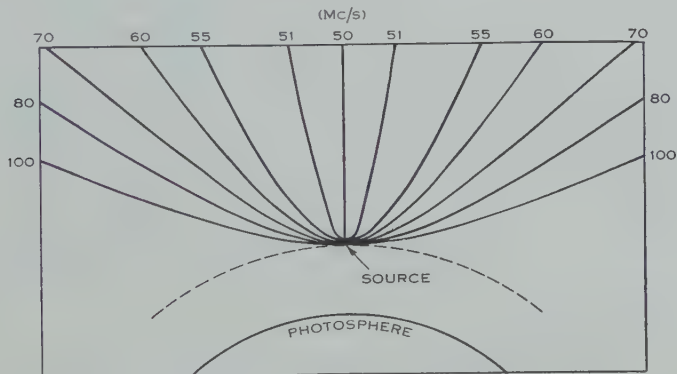


Fig. 7.—The limiting rays of outward emission at various frequencies from a source at the 50 Mc/s plasma level in the Baumbach-Allen corona. The numbers refer to frequencies in Mc/s. Data from Jaeger and Westfold (1950).

(c) *Conclusions on the Mechanism of Burst Generation*

The general agreement between the above interpretation and the observations supports the basic proposition that the sources of sporadic bursts radiate their energy mainly at harmonics of frequencies near the plasma frequency of the surrounding medium. Indeed it appears that, of the processes which have been suggested to explain the high intensity of solar noise, those involving plasma oscillations are the only ones capable of accounting for the harmonic phenomenon. Nevertheless, it should be stressed that our knowledge of plasma oscillations is far from complete and the conditions under which electromagnetic radiation can be obtained from them are not properly understood.

Evidence will be given in Section V that the source regions are in rapid motion through the solar corona and it is likely that the observed movements are associated with streams of ionized matter. As Shklovsky (1946), Bohm and Gross (1949), and others have pointed out, ionized streams projected through the corona provide a means of exciting plasma oscillations. Oscillations of this

kind are longitudinal and, if the amplitude becomes appreciable in comparison with the wavelength, the charge density exhibits excess "bunching" near the nodal points. This effect may provide the required non-linearity and asymmetry for the generation of the second harmonic.

V. THE POSITION, MOVEMENT, AND SIZE OF THE SOURCES OF BURSTS

(a) *The Position and Speed of the Sources of Type II Outbursts*

The conclusions of Section IV (b) suggest that in idealized circumstances it should be possible to determine both the source height and the source angle as a function of time directly from the dynamic spectrum of a burst. It is of course necessary to assume a standard electron-density model of the corona, and it is only possible to treat simple cases in which the observed profiles resemble the theoretical ones (Fig. 6 (c)). Under these conditions we can determine (1) the plasma frequency (and hence the height in the corona) from the observed peak frequency of the second harmonic, and (2) the critical escape frequency (and hence the source angle) from the observed cut-off frequency of the fundamental.

The relation between the critical escape frequency f_c , the plasma frequency f_0 , and the source angle θ is a complex one, but it has been shown by Smerd (unpublished data) that for the Baumbach-Allen corona the approximate formula

$$f_c = f_0 \sec (0.87\theta)$$

agrees with the values calculated by Jaeger and Westfold (1950) to within a few per cent. for frequencies between 20 and 100 Mc/s and source angles between 0 and 80°. A tabulation of plasma frequency for various coronal heights has been given by Smerd (1950).

The application of this method of source location is demonstrated in Figure 8 for the outburst shown in Figure 3. Figure 8 (a) shows plots of the observed values corresponding to f_0 and f_c ; Figure 8 (b) the derived source height and angle as a function of time; and Figure 8 (c) successive positions of the source at $\frac{1}{2}$ -min intervals. The calculated path is seen to be mainly across the line of sight. Since the active areas on the disk at the time of the outburst were confined almost entirely to western heliographic longitudes (see Fig. 8 (c)) we might infer that the source travelled across the disk from west to east. The mean speed of travel is found to be about 4000 km/sec.

The results given in Figure 8 emphasize the importance of knowing the rate of change of source angle. In previous determinations of the velocity of outburst sources, either the determination was confined to the radial component (spectroscopic method, Wild 1950a) or the speed was calculated on the assumption that the motion was strictly radial (directional method, Payne-Scott and Little 1952). Had the speed been derived on the assumption of radial motion in the present case, the less plausible conclusion would have been reached that the source initially travelled slowly inwards and subsequently turned and accelerated outwards.

The outburst of Figure 2 is less suitable for similar analysis partly because of its greater complexity and partly because the profile data are less accurate. Near the finish, however, the peaks are well defined, and the closeness of the

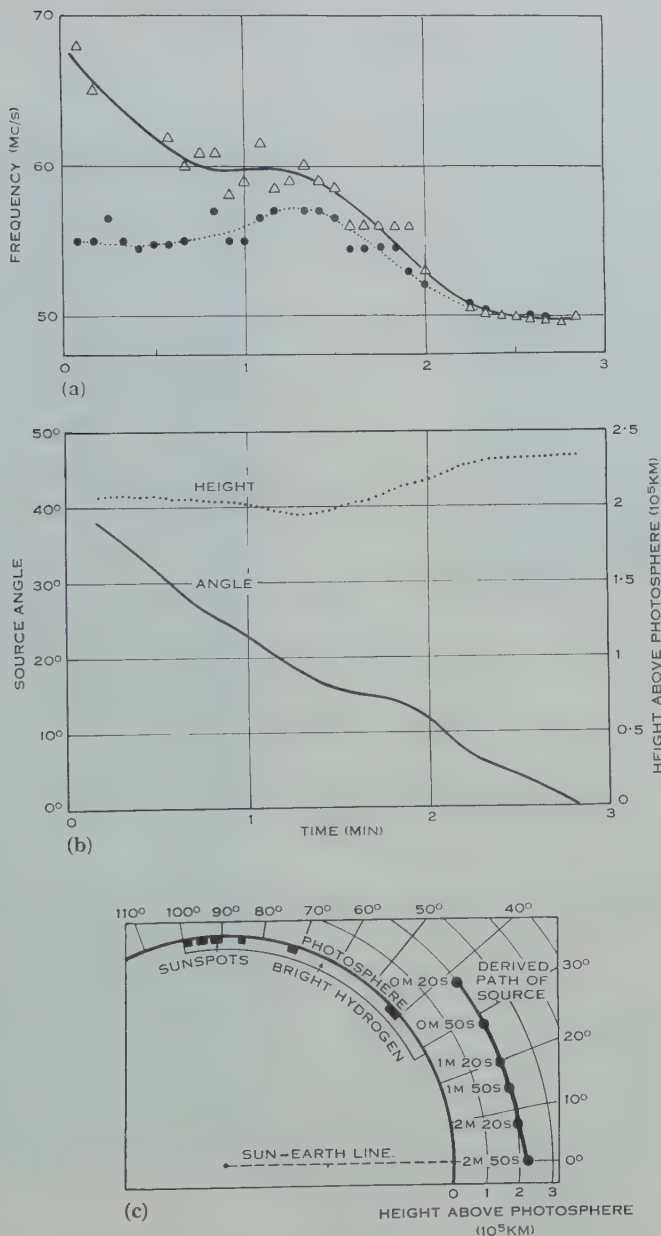


Fig. 8.—Position determinations for the outburst in Figure 3. (a) The variation with time of (i) the peak frequency of the fundamental (triangles) and (ii) half the peak frequency of the second harmonic (dots); (b) The derived height and "source angle" (θ in Fig. 6); (c) Successive positions of the source at $\frac{1}{2}$ -min intervals. The heliographic longitudes of sunspots etc. are indicated, and refer to the western hemisphere.

peak-frequency ratio to 2 suggests a central position on the disk. Since the associated flare was also located near the centre, it may be inferred that the source travelled approximately radially outwards in the direction of the Earth. The derived height plot is shown in Figure 9, and represents the most extensive range of travel we have yet observed in an outburst. The plot indicates an approximately constant velocity of 475 km/sec over a 6:1 range in heights. The time delay between the fadeout and the start of the outburst suggests that the source could have been ejected simultaneously with the onset of the ultra-violet emission from a region low in the corona, some 2.5×10^4 km above the photosphere.

The emission of second harmonics in outbursts suggests a reason for the unexpectedly large source heights deduced by Payne-Scott and Little (1952) from their directional observations. These authors considered the typical

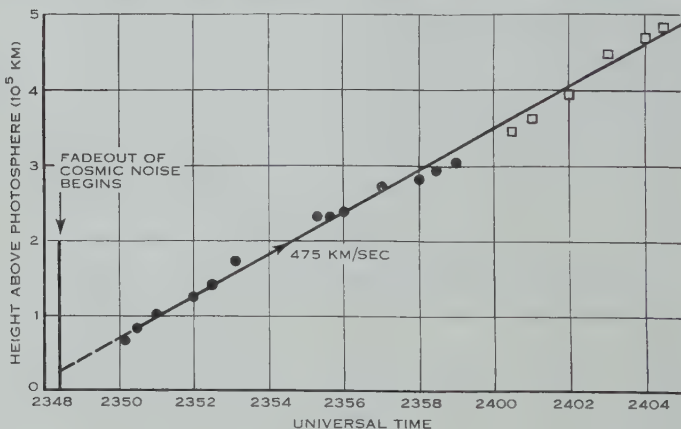


Fig. 9.—Derived height plot for the outburst in Figure 2. Where possible heights are derived from the fundamental frequency band (dots). The range is extended to greater heights by using half the frequency of the second harmonic (squares).

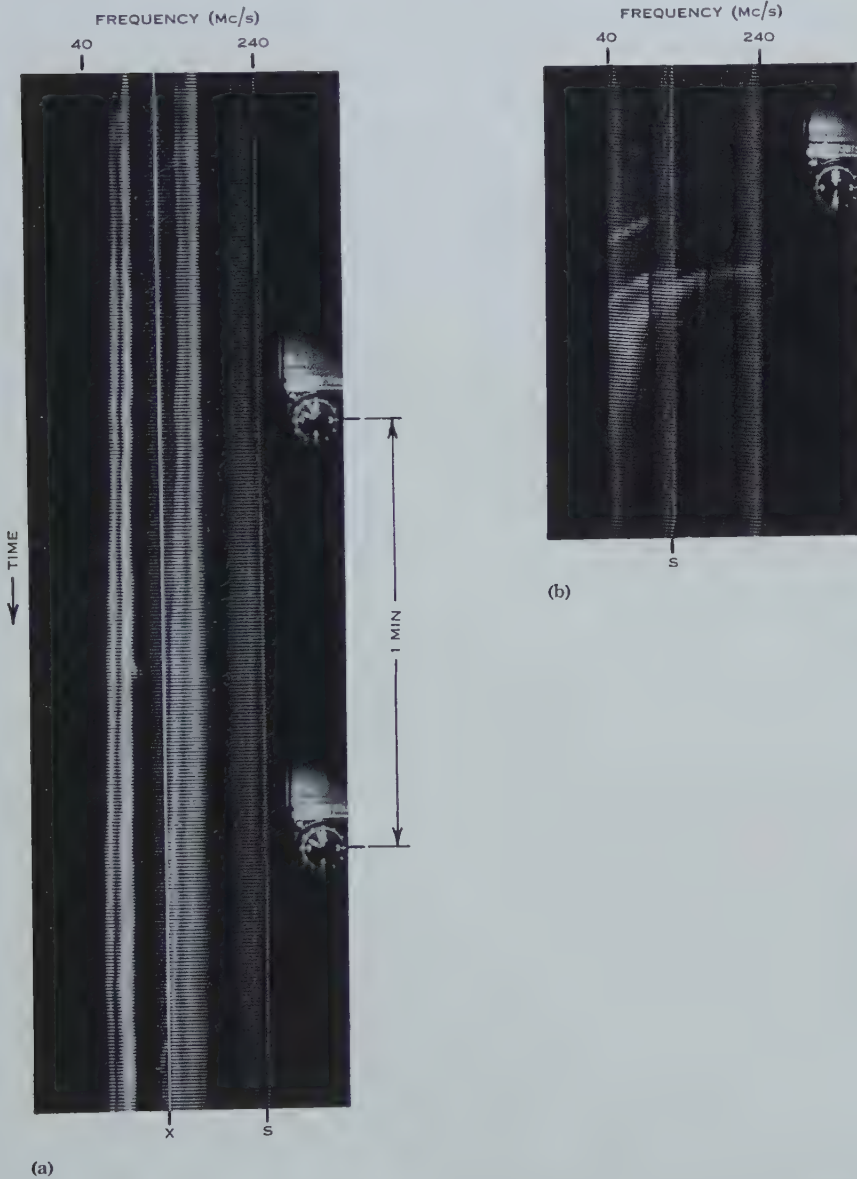
minimum height for 97 Mc/s radiation to be some 2×10^5 km above the photosphere. This, they point out, is about three times as high as the 97 Mc/s plasma level in the Baumbach-Allen corona. Rather than indicating a radical alteration to the electron density distribution, as suggested by these authors, the discrepancy could be simply accounted for in terms of emission at the second harmonic: for the assumed electron-density distribution, the height of the 97/2 Mc/s plasma level is 2.4×10^5 km.

(b) Sources of Error in Velocity Determinations

The two main sources of error in the above method of velocity determination are likely to be (i) the use of a particular spherically symmetrical model of the corona, and (ii) the neglect of coronal magnetic fields. We now consider these in turn.

(i) *Coronal Irregularities.*—It is well known that the electron-density distribution of the corona can depart considerably from the spherically sym-

HARMONICS IN THE SPECTRA OF SOLAR RADIO DISTURBANCES



Intensity-modulated records showing harmonics in (a) part of a type II burst (corresponding to Fig. 2) and (b) a type III burst (corresponding to Fig. 4 (f)). The constant frequency signals marked *S* are due to local transmitters. The vertical white line *X* is not part of the record.

metrical model we have assumed. In the first place the *mean* electron-density distribution changes with the solar cycle. From his analysis of optical data, van de Hulst (1950) concludes that the mean density at any height decreases by a factor of about 1.8 between maximum and minimum phases. For the equatorial region, his sunspot-maximum model agrees closely with the Allen-Baumbach model, while heights in his sunspot-minimum model are about 5×10^4 km lower in the range of interest here. However, the spacing between levels of different electron density is scarcely affected; for instance the spacing between the 50 and 100 Mc/s levels changes by only 5 per cent. This indicates that the determination of velocity along radial directions is not significantly affected by changes in the mean distribution with the solar cycle. Effects due to departures from the mean values, manifest in the complex structure of the corona, are likely to be much more important. The optical data, which have recently been reviewed by van de Hulst (1953), seem to indicate that the electron density in coronal rays may be some 10 times that between the rays while the gradients in the two regions are similar.

Let us now re-examine the outburst analysed in Figure 9, bearing in mind the effects of coronal structure. Assuming the true source velocity to be constant, the linearity and smoothness of the height plot indicate that the "shape" of the assumed electron-density law is the correct one along the path of the source. Suppose, however, that the source travelled along a coronal ray whose density was 10 times that given by the assumed model. Source heights are then increased by some 4×10^5 km, and the velocity (still essentially constant) by a factor of 1.6. Taking van de Hulst's sunspot-minimum model to represent the other extreme, it is estimated that the outburst's true outward component of velocity probably lay between 400 and 750 km/sec.

In cases where the motion is largely transverse (e.g. Fig. 8), the derivation of speed depends on the large-scale angular distribution of electron density, and the quantitative results are less certain.

(ii) *Coronal Magnetic Fields*.—The presence of a magnetic field at the source causes two additional proper frequencies to exist either side of the plasma frequency (Westfold 1949). The three frequencies correspond to those of the magneto-ionic theory at which the refractive index vanishes; each shows a characteristic polarization. At the coronal heights with which we are concerned the three frequencies are probably closely spaced (i.e. $f_H \ll f_0$) and the effects of neglecting magnetic fields could cause no significant errors in the derivation of velocities.

Application of the magneto-ionic theory shows that from any one point only two of the three frequencies can escape from the Sun. It was previously suggested (Wild 1950a) that the presence of closely spaced double peaks, like those in Figures 2 and 3, could be due to this effect of magnetic splitting. An alternative explanation is in terms of two sources separated in space. The former explanation now seems more likely because in both instances described here the two peaks are seen to fade simultaneously, thus suggesting a common source. Combined observations of spectrum and polarization should help to decide this question.

(c) The Size of Outburst Sources

In the simple interpretation outlined in Section IV it was assumed that burst radiation emanates from a point source. For sources of finite size we should receive the spectrum due to the excitation of a finite band of plasma frequencies. Lack of knowledge of the natural spectrum prohibits the estimation of source dimensions from the width of the observed peaks. We can, however, use the bandwidth data to set an upper limit on at least one of these dimensions.

In the vicinity of 50 Mc/s, bandwidths as small as 1.5 Mc/s between half-power points are not uncommon. With the standard coronal model this means that the source responsible for the peak is contained within a height range of less than 10^4 km. Also, in cases such as Figure 3, it can be inferred from the abruptness of the cut-off that the gradient of the outer edge of the source is such that the intensity may change by a factor of 2 within a height range of less than 500 km.

(d) The Sources of Type III Bursts

The main conclusions of Section IV, that the burst-generating sources probably radiate at harmonics of the plasma frequency of the surrounding medium, apply to type III as well as to type II bursts. This provides an experimental foundation to the speculation (Wild 1950b) that the fast frequency drift of type III bursts is to be interpreted in a similar fashion to the slow drift of type II bursts. On this basis the type III sources are found to move outwards through the solar corona with initial radial velocity components of between 3×10^4 and 10^5 km/sec, showing steady deceleration along their path. Those with "tails" are apparently brought to rest at heights of a few hundred thousand kilometres.

The interpretation of type III bursts has been discussed briefly by Wild, Roberts, and Murray (1954) and will be considered in more detail in a later paper.

VI. SUMMARY OF CONCLUSIONS

The observational results of this paper are summarized at the end of Section III. The presence in the spectra of sporadic bursts of intense narrow emission bands accompanied by their second harmonics indicates that the sources consist of oscillatory charges which oscillate, according to a non-linear law, at or near some proper frequency. From considerations of the escape of radiation from resonance levels in the solar atmosphere it is concluded that the plasma frequency is the only known proper frequency capable of accounting for the observations. If the natural spectrum of emission from the plasma level is assumed to show a slight spread in frequency about the plasma frequency, it is possible to give a natural explanation of certain peculiar features of observed spectra. The sources are found to be in rapid motion through the corona and it is suggested that the generation of high intensities is associated with longitudinal plasma oscillations excited by fast streams of ionized matter.

Assuming a standard, spherically symmetrical corona, application of the plasma hypothesis to observed spectra yields information on the position (both height and angular displacement from the centre of the disk), velocity, and size

of the source. Velocities of 500 and 4000 km/sec were deduced for two long-duration outbursts, and velocities as great as 10^5 km/sec for the short-lived type III bursts.

VII. ACKNOWLEDGMENTS

The authors wish to thank Dr. J. A. Roberts and Mr. S. F. Smerd for valuable discussions on theoretical aspects of this work, and Mr. J. Joisee for assistance in the regular running of equipment.

VIII. REFERENCES

- BOHM, D., and GROSS, E. P. (1949).—*Phys. Rev.* **75**: 1851, 1864.
VAN DE HULST, H. C. (1950).—*B.A.N.* **11**: 135.
VAN DE HULST, H. C. (1953).—"The Sun." Ch. 5. (Univ. Chicago Press.)
JAEGER, J. C., and WESTFOLD, K. C. (1949).—*Aust. J. Sci. Res.* **A 2**: 322-34.
JAEGER, J. C., and WESTFOLD, K. C. (1950).—*Aust. J. Sci. Res.* **A 3**: 376-86.
MARTYN, D. F. (1947).—*Nature* **159**: 26-7.
PAYNE-SCOTT, RUBY, and LITTLE, A. G. (1952).—*Aust. J. Sci. Res.* **A 5**: 32-49.
ROBERTS, J. A. (1952).—Incoherent radiation of electrons in a magnetic field. Ph.D. Thesis, University of Cambridge.
RYLE, M. (1948).—*Proc. Roy. Soc. A* **195**: 82-97.
SCHWINGER, J. (1949).—*Phys. Rev.* **75**: 1912.
SHKLOVSKY, I. S. (1946).—*Astr. J., Moscow* **23**: 333-47.
SMERD, S. F. (1950).—*Proc. Instn. Elect. Engrs.* **III 97**: 447-52.
WESTFOLD, K. C. (1949).—*Aust. J. Sci. Res.* **A 2**: 169-83.
WILD, J. P. (1950a).—*Aust. J. Sci. Res.* **A 3**: 399-408.
WILD, J. P. (1950b).—*Aust. J. Sci. Res.* **A 3**: 541-57.
WILD, J. P. (1951).—*Aust. J. Sci. Res.* **A 4**: 36-50.
WILD, J. P., and MCCREADY, L. L. (1950).—*Aust. J. Sci. Res.* **A 3**: 387-98.
WILD, J. P., MURRAY, J. D., and ROWE, W. C. (1953).—*Nature* **172**: 533-4.
WILD, J. P., ROBERTS, J. A., and MURRAY, J. D. (1954).—*Nature* **173**: 532.

OBSERVATIONS OF COSMIC NOISE AT 9.15 Mc/s

By C. S. HIGGINS* and C. A. SHAIN*

[*Manuscript received April 22, 1954*]

Summary

From observations made at a frequency of 9.15 Mc/s, with an aerial of beam width 29° between half-power points and directed to Dec. -32° , a curve of equivalent aerial temperature, as a function of sidereal time, is derived.

The temperatures observed were of the order of 10°K . The curve is compared with curves derived for similar conditions by calculation from the results of observations at 18.3 Mc/s and at 100 Mc/s. It is found that the equivalent temperatures increase rapidly with decreasing frequency, but the ratio of maximum to minimum temperature decreases with frequency.

It is shown that "atmospheric" noise levels observed by the standard techniques sometimes contain a large contribution from cosmic noise at this frequency.

I. INTRODUCTION

The distribution over the sky of the intensity of cosmic noise has been studied at a number of widely spaced frequencies, and attempts have been made to construct theoretical models of the Galaxy based on these radio observations. It has been found that such models depend, for the quantitative evaluation of important empirical constants, on observations at comparatively low frequencies, about 10 or 20 Mc/s.

Observations in this range of frequencies are rare, the only published work at a frequency close to 10 Mc/s consisting of a few measurements at 9.5 Mc/s by Friis and Feldman (1937) which were made during tests of the original MUSA aerial. A recent paper (Shain and Higgins 1954) presented the results of a detailed survey of a restricted region of the sky at 18.3 Mc/s, but the results of some earlier work at the same frequency (Shain 1951), in which a strip of the sky was scanned by a fixed aerial directed to a constant declination, have already been used by Piddington (1951) and Brown and Hazard (1953) for comparison with their theoretically predicted intensities. Observations with such a fixed aerial are much simpler to make than a general survey and, since equipment was available which could be readily adapted for the purpose, an attempt was made to obtain similar observations at a frequency of 9.15 Mc/s. The present paper describes the results of these observations.

In the course of the present observations, interference was experienced from atmospherics and from radio stations, but, at times when ionospheric absorption was small, a background intensity, which varied with sidereal time, was observed consistently and this was undoubtedly due to cosmic noise.

* Division of Radiophysics, C.S.I.R.O., University Grounds, Sydney.

The variations in the intensity of cosmic noise at 9.15 Mc/s revealed in the present work cannot be directly compared in detail with the variations observed at other frequencies since different aerial systems have been used. Nevertheless a comparison is made between the observed variations at 9.15 Mc/s and those to be expected, by calculation from published observations, if aerials of similar directivity had been used at 18.3 Mc/s and at 100 Mc/s. The comparison shows that the trends with frequency expected from work at higher frequencies continue down to 9 Mc/s.

The intensity of cosmic noise at 9.15 Mc/s is high, equivalent temperatures being of the order of 10^6 °K. The observation of such high intensities suggests that some of the measurements of noise levels which have been assumed to be due to atmospherics might, in fact, have referred to cosmic noise, and this point will be briefly discussed.

II. EQUIPMENT

(a) Aerial

The aerial was erected at Hornsby, N.S.W. (lat. 33.7° S., long. 151.1° E.). It consisted of an array of 12 horizontal half-wave dipoles, 0.1 wavelength above the ground, arranged in plan as shown in Figure 1 (a). To reduce ground losses, wires parallel to the dipoles and extending $\frac{1}{4}$ wavelength beyond each

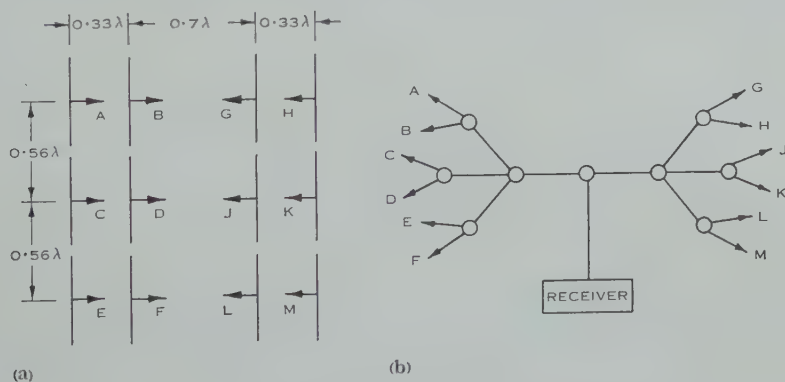


Fig. 1 (a).—Plan of the array.

Fig. 1 (b).—Connection of feeders. Each circle represents a T-type matching network. Letters indicate feeders to dipoles as shown in (a).

end of the array were laid 2 ft apart on the ground underneath the array. Each dipole had a balance-unbalance transformer and a separate coaxial feeder which was taken to a hut in the centre of the array. In this hut the feeders were grouped as shown in Figure 1 (b) with a matching network at each junction to match the impedances to the characteristic impedance of the cables used. The receiving equipment was housed in another hut outside the array.

The calculated aerial diagrams for the north-south and east-west planes are shown in Figure 2; it is seen that the beam widths to half power are 31° and 26° in the north-south and east-west planes respectively. Owing to a slight slope of the ground downwards towards the north, the direction of maximum sensitivity

of the aerial was not overhead; the aerial scanned a strip of the sky centred on Dec. -32° . This strip included the galactic centre and the south galactic pole.

Losses in the ground reflecting system were estimated from impedance measurements on individual dipoles. It was assumed that the differences between the measured dipole impedances and the input impedances of similar dipoles in free space was due to the mutual impedance of the dipole and its image. The free space impedance and the mutual impedance of a dipole and its image were taken from data given by Schelkunoff and Friis (1952). From these calculations it was estimated that ground losses, compared with a perfectly reflecting ground, amounted to 2.3 db. The losses in the feeders and matching networks were measured directly and were found to be 2.2 db. Thus the total correction due to losses in the aerial system amounted to 4.5 db, corresponding to a factor of 2.8 with an estimated probable error in this factor of ± 20 per cent.

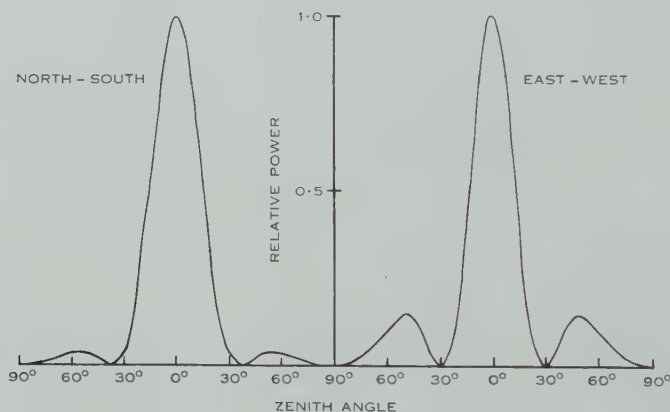


Fig. 2.—Calculated aerial sensitivity patterns.

(b) *Receiving and Recording Equipment*

A standard communications receiver was used, operated with a bandwidth of 400 c/s. Since the received noise powers were always high compared with the receiver noise, no attempt was made to achieve a low noise factor.

The output of the receiver was generally displayed on a recording meter. However, on some occasions, as described later, the time constant of the meter was too long and an alternative cathode-ray tube display was used with a time constant of less than 1 sec. In each case the record was calibrated using a diode noise generator the output impedance of which was made equal to the aerial impedance seen by the receiver.

III. OBSERVATIONS

The observations discussed in the present paper were made at intervals during the period July 1951 to September 1952. From previous experience at 18.3 Mc/s it was expected that often, and always during the day, ionospheric absorption would be high, and also that some difficulty would be experienced from atmospherics and from station interference. The objective, therefore,

was to obtain as large a number of records as possible, perhaps of only an hour's duration, during which these effects were not serious. In the course of a year sufficient samples to cover a full sidereal day could be obtained.

During the morning and early afternoon on most days the record had the appearance of random noise only and there was little interference either from atmospherics or from radio stations. However, as expected, the noise powers recorded at these times were always low, due to ionospheric absorption. In the late afternoon and early evening, especially in summer, atmospherics increased in intensity. These were recognized as "crashes" when listening to the receiver output through a loudspeaker, and as "spikes" on the record. However, provided the individual spikes were not too frequent, the lower edge of the trace showed a fairly well-defined base level of apparently random noise. During the same part of the day, station interference often became serious; as far as possible this was avoided by slight retuning of the receiver, although the frequency was always kept within 10 kc/s of 9.15 Mc/s. Atmospherics generally

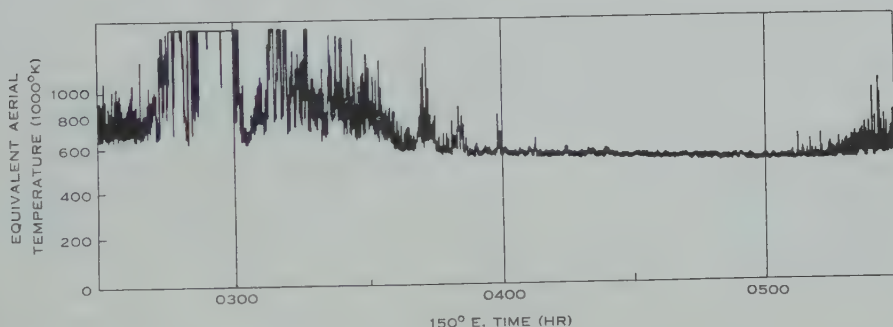


Fig. 3.—A record showing the absence of atmospherics and station interference for about an hour during the early morning of August 23, 1951. Sunrise occurred at 0627 hr (150° E. time). Atmospherics, which appear on the record as spikes, are very frequent at about 0230 hr. Interference from a station is severe from 0240 hr but decreases in intensity, together with the atmospherics, until the record is clear of interference by 0415 hr. Atmospherics increase in intensity again soon after 0500 hr. The sidereal time at 0400 hr (150° E. time) is 02 hr 05 min, and the gradual drop in received power is a real variation in cosmic noise intensity.

increased in intensity and frequency until about midnight, and commonly such periods of the records could not be used, although on a few occasions the record remained quite smooth throughout the night with no noticeable trace of atmospherics. Between midnight and sunrise, at times when the critical frequency of the *F* region was low, atmospherics and station interference decreased greatly in intensity. The majority of the useful observations were obtained during this period of the day. Figure 3 shows a record during which atmospherics and station interference decrease in intensity until the record is smooth from about 0415 to 0500 hr. Later, atmospherics increase again in intensity. It will be seen that even when the atmospherics spikes are moderately frequent the general variation of the cosmic noise base level can be followed. The fact that spikes due to atmospherics decreased, and on many occasions completely

faded out, at the same time as interference from radio stations, suggests that even atmospherics from distant lightning flashes do not average out sufficiently to give a smooth record. If this is so, absence of spikes could be taken to indicate absence of atmospheric interference.

During summer it was difficult to find records completely free of atmospherics, even during the early morning. Although their intensity was small, the frequency of the spikes was sufficient to render uncertain the cosmic noise base level when the recording meter was used. To overcome this difficulty, the receiver output was displayed on the cathode-ray tube with a time base duration of about 30 sec, and each trace was photographed separately on 16-mm film. After several frames the noise generator was substituted for the aerial and several frames were exposed at each of several known noise levels. It

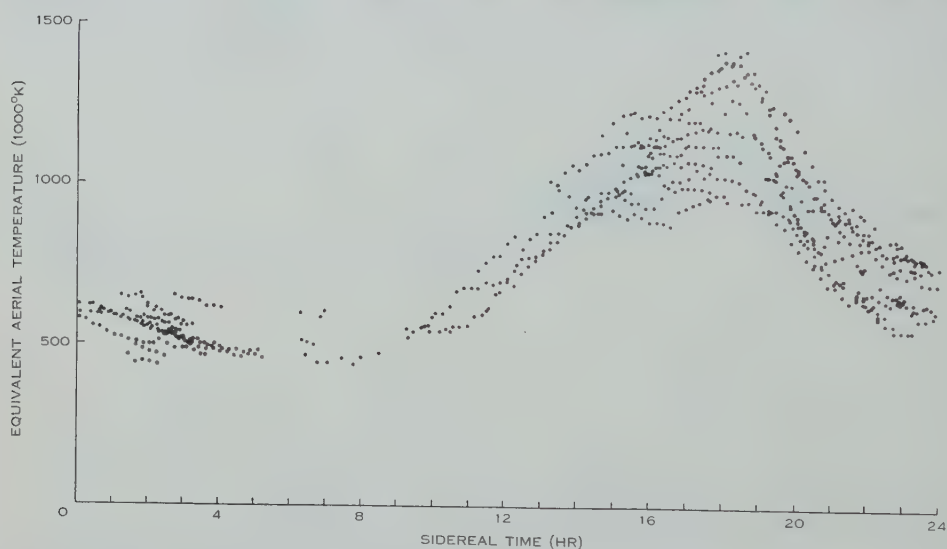


Fig. 4.—Observed equivalent aerial temperatures *v.* sidereal time (corrected for aerial losses).

was found that, on the occasions when this method of recording was used, the atmospherics could always be distinguished and a cosmic noise base level determined. When the intensity was later determined by observation of the film, it was found that, with visual integration of the base level over several minutes, the accuracy was no less than that obtainable with good meter records. However, the inconvenience of this technique accounts for the paucity of observations in the period between 05 and 09 hr, the region which passed through the aerial beam during early morning hours in summer time.

Readings were taken from all suitable records for periods between half an hour after sunset and half an hour before sunrise. When a record showed no interference for some time, readings were taken every 12 min. The results of all these observations, in the form of equivalent aerial temperatures corrected for losses in the aerial system, are plotted against the sidereal time of observation in Figure 4. Although a contribution from atmospherics to the received noise

powers plotted in Figure 4 cannot be completely ruled out, it is considered that any such contribution must be very small. This is confirmed by the fact that there is a marked sidereal diurnal variation of intensity in Figure 4, while intensities measured at roughly the same sidereal time at intervals of up to 4 months showed no systematic variation with season.

IV. ANALYSIS

(a) *Corrections for Ionospheric Absorption*

Since the observations were limited to the night hours, and mainly to the period between midnight and half an hour before sunrise, absorption in the *D* region could be neglected. However, Mitra and Shain (1953) have shown that appreciable absorption of cosmic noise takes place in the *F* region when the critical frequency of this region, f_oF_2 , is greater than about one-third of the operating frequency. There were insufficient observations to obtain a curve of absorption of 9.15 Mc/s cosmic noise against f_oF_2 and it was decided to use the results of the 18.3 Mc/s absorption measurements, assuming that *F*-region absorption depended only on the ratio of the operating frequency to f_oF_2 . For example, Mitra and Shain found that absorption at 18.3 Mc/s, measured at Hornsby, was 0.4 db when f_oF_2 at Canberra was 8 Mc/s; it was assumed that at 9.15 Mc/s the absorption would be 0.4 db when f_oF_2 was 4 Mc/s.

Values of f_oF_2 were taken from data issued by the Ionospheric Prediction Service of the Commonwealth Observatory in "Ionospheric Predictions—Series *D*". Where possible, the data for Canberra (lat. 35.3 °S., long. 149.0 °E.—about 250 km south-west of Hornsby) were used. Unfortunately, for observations during several months no Canberra data were available. At such times f_oF_2 at Canberra was estimated using the values observed at Brisbane (lat. 27.5 °S., long. 153.0 °E.—about 600 km north of Hornsby), and the trend of f_oF_2 with latitude indicated in "Ionospheric Predictions—Series *W*", also issued by the Ionospheric Prediction Service. Although this procedure introduces some uncertainty in the application of the corrections for ionospheric attenuation, it is not serious since in any case the corrections were all small, generally less than 10 per cent. Also attenuations calculated according to the above procedure using Brisbane data agreed well with values calculated directly from Canberra data when the latter were available.

The values of equivalent temperature shown in Figure 4, corrected where necessary for ionospheric attenuation, are replotted in Figure 5. The main part of the scatter of the points is due to the uncertainty of reading from the records. A smooth curve is drawn through the averages of all points within hourly intervals. The accuracy of the curve in Figure 5 is estimated to be better than 10 per cent. for relative values, except for the dashed section where it is somewhat less, and 20 per cent. for the absolute scale.

(b) *Comparison with Observations at Other Frequencies*

The curve in Figure 5 has features similar to those of corresponding curves for other frequencies. There is a pronounced maximum (near 18 hr) as the galactic centre passes through the aerial beam, and the minimum near 05 hr

agrees with the positions found at 18.3 Mc/s and at 100 Mc/s. The values of equivalent aerial temperature are, however, considerably higher. Direct comparison with observations at other frequencies is not possible since the temperatures observed, especially near the maximum, would depend fairly critically on the aerial directivity. In a previous paper (Shain 1954) it was pointed out that the aerals used in surveys by Shain and Higgins (1954) at 18.3 Mc/s and by Bolton and Westfold (1950) at 100 Mc/s had main lobes of nearly the same shape, although the side lobe pattern was different in each case. In addition, Shain corrected the results of these surveys for the effects of side lobes. This gave contours of equivalent aerial temperature which would have been obtained if, at each frequency, an idealized aerial which had a single lobe, the shape of the main lobe in the two surveys, had been used.

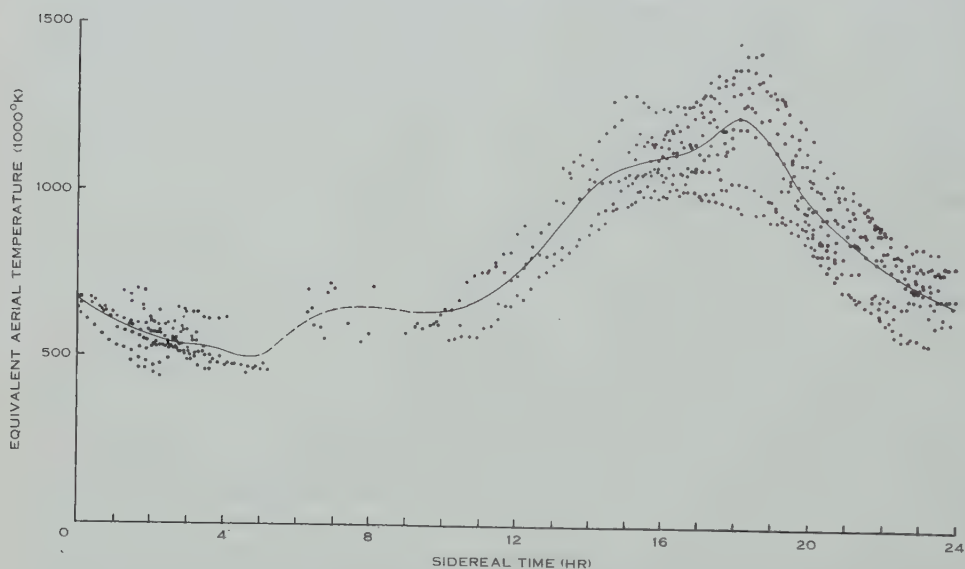


Fig. 5.—The observed equivalent aerial temperatures of Figure 4, corrected for the effects of ionospheric absorption. The curve is drawn through hourly averages, except for the dashed section between 4 and 9 hr.

The equivalent aerial temperature is the average, weighted according to the aerial sensitivity in different directions, of the brightness temperatures over the visible sky. If observed equivalent aerial temperatures are available for a single lobe aerial, aerial *A*, say, having a comparatively small beam width, it is possible to determine the equivalent temperature that would be observed by a broader-beamed aerial by taking suitably weighted averages of the equivalent temperatures seen by aerial *A* when pointed in appropriate directions. This procedure was adopted to obtain the equivalent aerial temperatures that would have been observed at 18.3 Mc/s and at 100 Mc/s with an aerial having the same directivity (including side lobes) and pointed in the same direction as that used for the 9.15 Mc/s observations.

Curves of the calculated equivalent temperature as a function of sidereal time for such hypothetical observations at 18.3 Mc/s and at 100 Mc/s are shown in Figure 6, together with the curve for observed 9.15 Mc/s equivalent temperatures taken from Figure 5. The scales of the curves have been adjusted so that the curves coincide near their minima.

Several points of interest in Figure 6 may be noted. The general shapes of the curves at the three frequencies are very similar, and the differences in shape show a progressive trend with frequency. The ratio of maximum to minimum temperatures decreases with frequency although the actual values of the equivalent temperatures increase rapidly as the frequency is lowered.

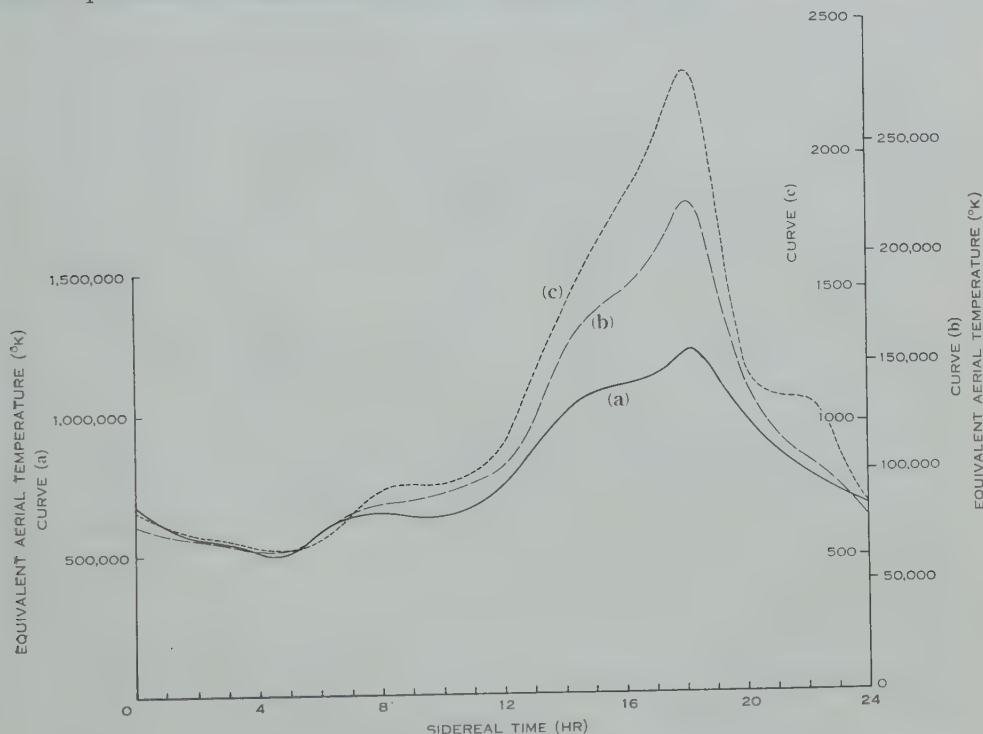


Fig. 6.—Comparison of equivalent temperatures at three frequencies. (a) Equivalent aerial temperature *v.* sidereal time as observed at 9.15 Mc/s (from Fig. 5). (b) Equivalent aerial temperature *v.* sidereal time to be expected at 18.3 Mc/s with an aerial similar to that used at 9.15 Mc/s; calculated from 18.3 Mc/s observations as explained in the text. (c) A similar curve to (b) but for 100 Mc/s.

Figure 7 shows the maximum and minimum equivalent temperatures and their ratio plotted against frequency using logarithmic scales. The points in Figure 7 (a) corresponding to the minimum temperatures are fitted closely by the relation.

$$T \propto f^{-2.8}$$

where T °K is the minimum equivalent temperature observed at a frequency f Mc/s.

The bump in each curve in Figure 6 centred on about 15 hr sidereal time is largely an instrumental effect, due to the side lobes of the aerial in the east-west plane, but it was found that, even after side lobe effects had been removed, small bumps were still evident on the curves for 18.3 and 9.15 Mc/s. Calculations showed that passage of the aerial beam over discrete sources previously observed at 18.3 Mc/s could account for these bumps at 18.3 Mc/s, and the presence of similar bumps at 9.15 Mc/s suggests that discrete sources are contributing appreciably to the radiation at 9.15 Mc/s. No accurate evaluation of source intensities can be made, but it appears that at 9.15 Mc/s they stand out against the background at least as clearly as at 18.3 Mc/s, possibly more so.

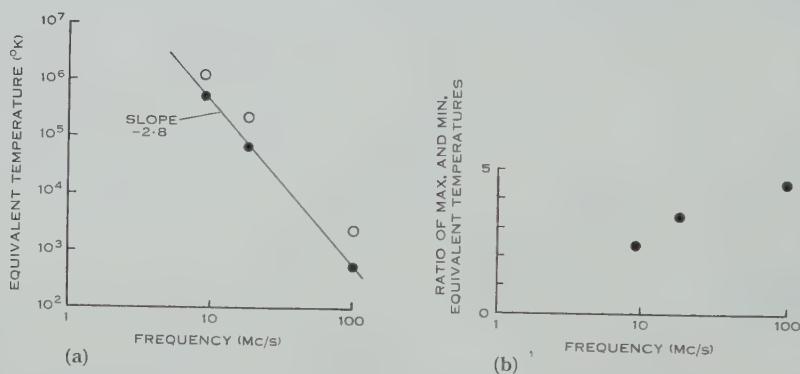


Fig. 7 (a).—Maximum (open circles) and minimum (full circles) equivalent temperatures (from Fig. 6) as a function of frequency.

Fig. 7 (b).—Ratio of maximum and minimum equivalent temperatures as a function of frequency.

V. DISCUSSION

It is not proposed to compare in detail the observational results with the predictions of published theories of the origin of the cosmic noise. Qualitatively, at least, the 9.15 Mc/s results fit the assumptions of an origin in discrete sources together with an absorbing (at 9.15 Mc/s) disk of interstellar gas. Shain (1954) showed that certain differences in the brightness distributions at 18.3 Mc/s and at 100 Mc/s could be accounted for if absorption in interstellar gas occurred at 18.3 Mc/s towards the galactic centre. From Shain's results, it would be expected that near the galactic centre the optical depth for 9.15 Mc/s radiation would be greater than unity for latitudes within 10° of the galactic equator. This could then account for the comparatively low value of the maximum temperature at 9.15 Mc/s and for the different variation with frequency of the maximum and minimum observed temperatures. It may be noted from Figure 7 that at 9.15 Mc/s the observed temperatures are still increasing rapidly with decreasing frequency. The rate of change of the minimum temperatures with frequency is greater (numerically) than that of the maximum temperatures. Observations at low frequencies may even show that minimum equivalent temperatures occur towards the galactic centre rather than near the pole.

The absolute values of the equivalent temperatures are considerably higher than values deduced from the observations of Friis and Feldman (1937). Townes (1947) has shown that the maximum noise intensity observed by Friis and Feldman at 9.5 Mc/s corresponded to an equivalent temperature of only 120,000 °K. Although their aerial was not directed near the galactic centre, their maximum value is considerably less than the minimum value obtained in the present investigation. Their aerial had its maximum sensitivity at low angles and consideration of the ionospheric conditions at the time of their observations (Gilliland *et al.* 1937) indicates that the low values recorded by Friis and Feldman were due to ionospheric attenuation.

Since Friis and Feldman could measure cosmic noise with a communications-type aerial, it is of interest to compare received cosmic noise powers with the noise field strengths measured during the world-wide survey described by Horner (1953). These noise levels were attributed to atmospherics. After considering the effect of receiver noise on the measurements, especially during daylight when measured noise levels are low, due to absorption, Horner also briefly considered cosmic noise. He concluded that under the normal conditions of measurement its intensity was too low to be significant by comparison with receiver noise.

Using a relation derived by Pawsey, McCready, and Gardner (1951), we find that the typical value of equivalent aerial temperature due to cosmic noise of 10^6 °K corresponds to a noise field strength of 4 db below 1 μ V/m for 10 kc/s bandwidth at 9.15 Mc/s. As an example of the noise levels measured in Horner's survey, the median noise levels at midnight throughout the year at Tatsfield, England, at 10 Mc/s, were within a few decibels of 10 db below 1 μ V/m for 10 kc/s bandwidth (i.e. some 6 db below the typical cosmic noise level). From September to February the medians were even lower than 10 db below 1 μ V/m. But these figures cannot be compared directly with the observed cosmic noise field strengths, since in Horner's survey vertical aerials, having maximum sensitivity at low angles, were used. Cosmic noise arriving at low angles will normally suffer considerable attenuation, and for large angles of incidence on the ionosphere it may be cut off completely. Under these conditions the cosmic noise levels measured with the vertical aerials would be much lower than the value given above. However, during the winter months at Slough, the mean midnight values of f_0F_2 were less than 3 Mc/s (Department of Scientific and Industrial Research 1953) so that at these times cosmic noise should not have been attenuated greatly, while the reception of atmospherics at 10 Mc/s would be poor. It therefore appears that, at least near midnight during winter, the measured noise level of 10 db below 1 μ V/m was largely due to cosmic noise. In summer, of course, f_0F_2 was higher and atmospherics probably predominated.

This investigation will not be pursued in this paper, but it is apparent that for frequencies of about 10 Mc/s cosmic noise, at least at some times, makes a major contribution to the noise levels observed during the "atmospheric noise level" programme.

VI. CONCLUSIONS

Results have been presented of observations of cosmic noise at 9.15 Mc/s made with a fixed aerial which scanned a strip of the sky centred on Dec. -32° . Although, if possible, a detailed survey with a sharper aerial beam should be undertaken, the observations described in this paper should prove useful in checking theoretical models of the radio Galaxy. With the smaller ratio of maximum to minimum temperatures and the large magnitude of the equivalent aerial temperatures, they continue trends established from observations at higher frequencies and, at least qualitatively, they are in accordance with what would be expected on current theoretical ideas.

The high intensity of cosmic noise suggests that it will be necessary, at some times, to allow for a considerable cosmic noise contribution in measurements of atmospheric noise levels.

VII. REFERENCES

- BOLTON, J. G., and WESTFOLD, K. C. (1950).—*Aust. J. Sci. Res. A* **3**: 19-33.
 BROWN, R. H., and HAZARD, C. (1953).—*Phil. Mag.* **44**: 939-63.
 DEPARTMENT OF SCIENTIFIC AND INDUSTRIAL RESEARCH (1953).—Characteristics of the ionosphere observed in Great Britain 1930-1946. Spec. Rep. Radio Res., Lond. No. 23. (H.M.S.O.: London.)
 FRIIS, H. T., and FELDMAN, C. B. (1937).—*Proc. Inst. Radio Engrs., N.Y.* **25**: 841-917.
 GILLILAND, T. R., KIRBY, S. S., SMITH, N., and REYMER, S. E. (1937).—*Proc. Inst. Radio Engrs., N.Y.* **25**: 823-40.
 HORNER, F. (1953).—Measurements of atmospheric noise at high frequencies. Spec. Rep. Radio Res., Lond. No. 26. (H.M.S.O.: London.)
 MITRA, A. P., and SHAIN, C. A. (1953).—*J. Atmos. Terr. Phys.* **4**: 204-18.
 PAWSEY, J. L., MCCREADY, L. L., and GARDNER, F. F. (1951).—*J. Atmos. Terr. Phys.* **1**: 261-77.
 PIDDINGTON, J. H. (1951).—*Mon. Not. R. Astr. Soc.* **111**: 45-63.
 SCHELKUNOFF, S. A., and FRIIS, H. T. (1952).—"Antennas—Theory and Practice." (John Wiley & Sons, Inc.: New York.)
 SHAIN, C. A. (1951).—*Aust. J. Sci. Res. A* **4**: 258-67.
 SHAIN, C. A. (1954).—*Aust. J. Phys.* **7**: 150-64.
 SHAIN, C. A., and HIGGINS, C. S. (1954).—*Aust. J. Phys.* **7**: 130-49.
 TOWNES, C. H. (1947).—*Astrophys. J.* **105**: 235-40.

THE ASSOCIATION OF PULSATING AND FLAMING AURORAS WITH COMPLETE IONOSPHERIC ABSORPTION AT MACQUARIE ISLAND

By G. MAJOR*

[*Manuscript received March 2, 1954*]

Summary

Records of simultaneous auroral observations and ionosphere soundings at Macquarie Island (geomagnetic lat. -61.7°) show that pulsating or flaming auroras are frequently accompanied by complete absorption of the vertically incident waves. However, the nocturnal variations of frequency of occurrence of these two phenomena are markedly different in form.

I. OBSERVATIONS

Heppner, Byrne, and Belon (1952) have reported that at College, Alaska, a pulsating aurora at the zenith is frequently associated with complete absorption of vertical incidence radio waves. Pulsating and flaming auroral displays were observed at the Australian National Antarctic Research Expedition Station at Macquarie Island (lat. $54^\circ 30' \text{ S.}$, long. $158^\circ 57' \text{ E.}$, geomagnetic lat. -61.7° , magnetic dip 78°) on 48 nights between August 14, 1950 and April 15, 1951 (Parsons and Fenton 1953), and on 36 of these nights simultaneous ionospheric records were obtained with the vertical incidence automatic variable frequency recorder (Cohen 1952; Jeffrey 1953). Ionospheric soundings which took 2 min for the sweep from 1 to 13 Mc/s were made every 10 min, and all these records have been rescaled and matched with the corresponding auroral observations. Auroral observations were made sporadically and time interpolation of auroral observations to coincide with ionosphere recorder timing was often necessary. However, interpolation was considered valid only when the pulsating or flaming state was constant and interpolation was never made over a period greater than 5 min. The ionosphere recorder programme was controlled by a synchronous motor run from the Macquarie Island Station power mains and, owing to unavoidable power frequency variations, the times of recording are not as reliable as those appertaining to the auroral observations. The errors in simultaneity are not considered to be appreciable and their effect on the results should be negligible.

The auroral observers reported (Parsons and Fenton loc. cit.): "Pulsating and flaming auroras were usually associated with very intense displays. Almost invariably confined to the northern sky or the region close to the zenith, they usually took the form of diffuse surfaces or scattered remnants of previously brighter draperies or coronas. The intensity was seldom very great. On no occasion did pulsating or flaming forms appear until a display had been in

* Antarctic Division, Department of External Affairs, Melbourne.

progress for some considerable time and on several occasions such forms were observed to persist until dawn.

Flaming auroras exhibited a general tendency for an upward sweep towards the magnetic zenith. Regular wave-like variations were infrequent but were of two distinct types, one in the nature of waves which brightened scattered areas of glow momentarily as they passed, the other in the form of a succession of regular arcs sweeping rapidly in a direction normal to their orientation. These latter were almost invariably directed towards the magnetic zenith but on one occasion very distinct waves following each other at intervals of about one second swept from approximately 20° south of the zenith to low in the northern sky."

II. RESULTS

(a) Pulsating and Flaming Auroras Associated with Complete Absorption

Table 1 shows the results of an analysis of all reported pulsating or flaming auroral forms with simultaneous ionosphere soundings. On not one occasion were F echoes present without E_s and on the great majority of occasions when E_s was present F echoes were blanketed. An aurora in a region up to 10° in radius centred on the geographical zenith is classed as a zenith aurora.

TABLE 1
OCCURRENCE OF COMPLETE ABSORPTION DURING SPECIFIC AURORAL DISPLAYS*

Condition of the Sky	Total No. of Ionospheric Traces	Percentage of Traces showing Complete Absorption
Pulsating or flaming aurora	182	62
Zenith pulsating aurora ..	80	61
Non-zenith pulsating aurora	53	60
Zenith flaming aurora ..	17	65
Non-zenith flaming aurora ..	31	65

* The average percentage occurrence of absorption for all occasions irrespective of auroral appearance is 18 per cent.

(b) Time and Seasonal Variations

All the hourly records of the sporadic E -region between August 1950 and April 1951 inclusive during the hours when auroras could have been seen if the meteorological conditions were favourable were analysed to determine the percentage of observations when complete absorption was present. (The maximum amount of twilight in which an aurora may be seen is somewhat subjective and therefore the time limits when an aurora was seen by Parsons and Fenton were taken as a guide—roughly the period is when the Sun is lower than 12° below the horizon.) Complete absorption was seen to vary with hour and with month. The dashed line in Figure 1 shows the variation of percentage

of observations showing complete absorption with time of day for all days, independent of their magnetic character.

The magnetic character of the days on which simultaneous pulsating and flaming aurora or both and ionosphere records were made is shown in Table 2.

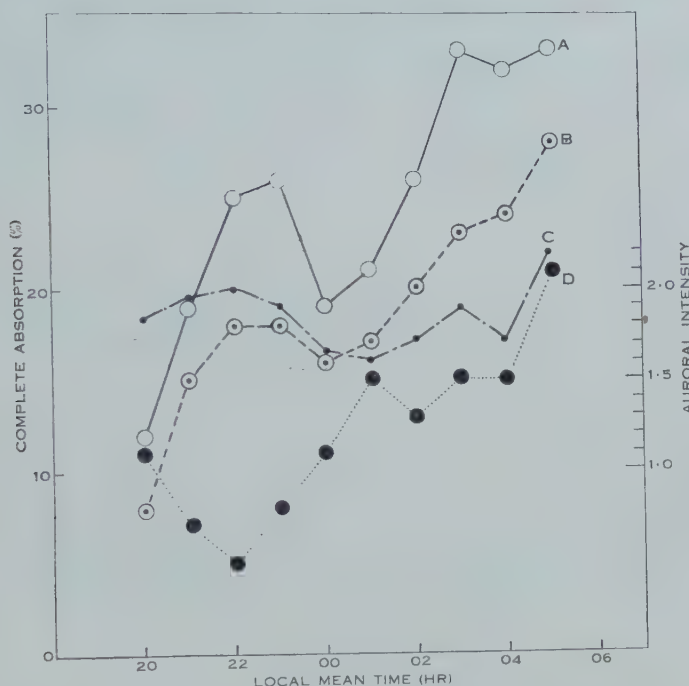


Fig. 1.—Curve A, nocturnal variation of percentage of observations showing complete absorption for the period August 1950 to April 1951 at Macquarie Island for all days except the 10 magnetically quiet days; curve B, as curve A except that it includes all days of each month; curve C, nocturnal variation of auroral intensity at Macquarie Island for the period May 1950 to April 1951; curve D, as curve B for College, Alaska for period September 1950 to April 1951.

Hourly E_s records for days other than the 10 quiet days of each month have been analysed revealing complete absorption characteristics as shown by the full line curve of Figure 1.

TABLE 2
MAGNETIC CHARACTER OF DAYS ON WHICH PULSATING AND
FLAMING OR BOTH AURORAS OCCURRED

Magnetic Character	No. of Observations
5 disturbed days of the month ..	9
5 quiet days of the month ..	1
10 quiet days of the month ..	3
"Normal" days	24

For comparison, the dotted line curve of Figure 1 shows the percentage of observations showing complete absorption at College, Alaska, for the hours between the end of civil twilight and the beginning of civil twilight for every day of the period September 8, 1950 to April 16, 1951. Also for comparison is the

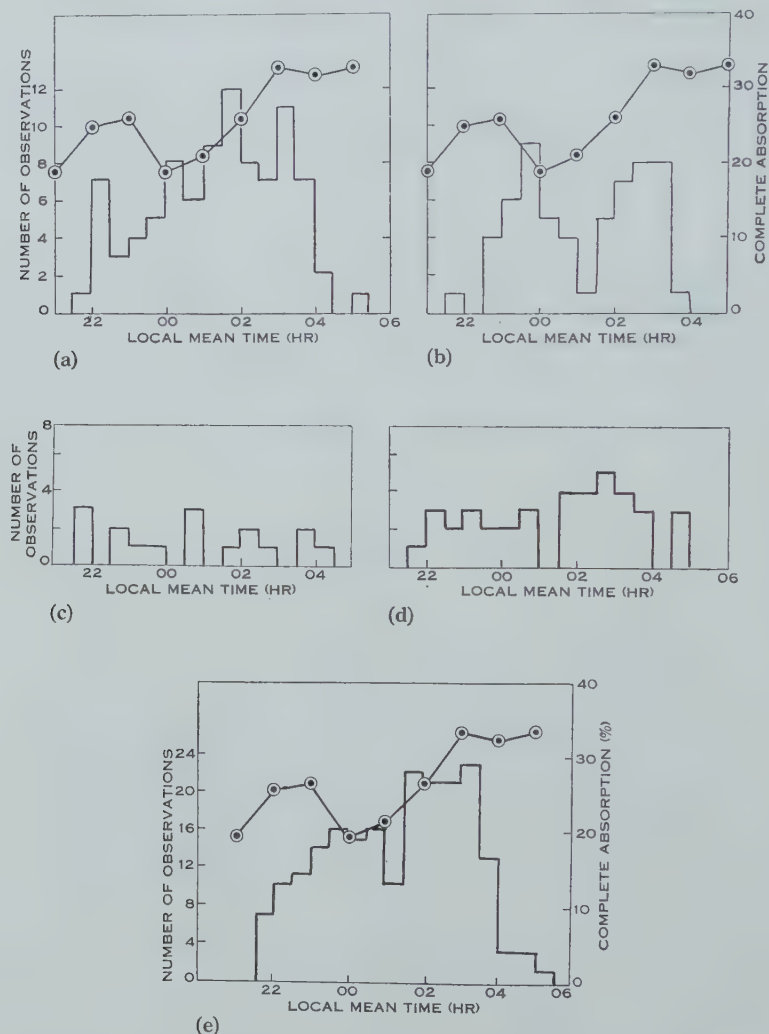


Fig. 2.—Nocturnal variation of observations of zenith and non-zenith pulsating and flaming auroras at Macquarie Island during the period August 1950 to April 1951. Curve A of Figure 1 is superimposed for comparison. (a) Zenith pulsating auroras; (b) non-zenith pulsating auroras; (c) zenith flaming auroras; (d) non-zenith flaming auroras; (e) all pulsating and flaming auroras.

broken line curve of Figure 1 (Jacka 1954) which shows the nocturnal variation of intensity of the aurora at Macquarie Island between May 1950 and April 1951. This curve gives the intensity of the brightest display (any form, any direction), observed during each hour, centred on the hour L.M.T.

The time of occurrence of pulsating and flaming forms is shown in Figures 2 (a)-(e) inclusive where the ordinates show the number of observations occurring within $\frac{1}{2}$ hr periods. Superimposed on Figures 2 (a), (b), and (e) is the full line curve of Figure 1.

Figure 3 shows the monthly variation of occurrence of pulsating and flaming auroras and of total absorption for both 1950-51 and 1951-52 during hours when auroral observations were possible. The pronounced equinoctial effect may be exaggerated because less hours were available for observation in December and January than in other months.

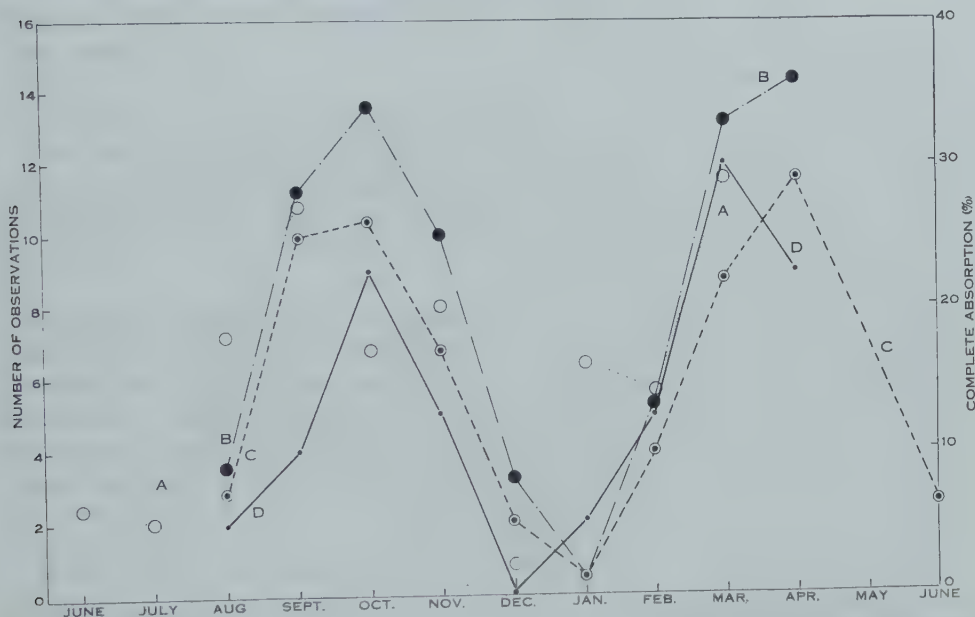


Fig. 3.—Curve A, seasonal variation of observations showing complete absorption during “auroral hours” for 1951-52 at Macquarie Island; curve B, as curve A except that it refers to all days except the 10 magnetically quiet days of each month for 1950-51; curve C, as curve B except that it refers to all days of each month; curve D, seasonal variation of observations of pulsating and flaming aurora at Macquarie Island for the period August 1950 to April 1951.

III. DISCUSSION OF RESULTS AND CONCLUSIONS

Table 1 and Figure 1 indicate that the percentage of traces showing complete absorption during pulsating and flaming auroral displays or both is at least twice the highest average percentage for any hour and may be compared with the results of Heppner, Byrne, and Belon (*loc. cit.*), who found 72 per cent. complete absorption from 239 traces during pulsating aurora at the zenith. The average percentage of traces showing complete absorption for all hours and all months at Macquarie Island is 18 per cent., while at College it is 14.5 per cent. It is interesting to note that the maximum frequency of occurrence of pulsating aurora near the zenith is at about 0200 L.M.T. at both locations. It is interesting also to note that non-zenithal and zenithal pulsating and flaming auroras give similar ionosphere absorption percentages at Macquarie Island.

The maximum of the full line curve of Figure 1 at about 2300 L.M.T. (which is also local mean magnetic midnight) and the minimum 1 hr later do not appear on any of the aurora occurrence diagrams of Figures 2 (a)-(e); indicating that neither this maximum nor minimum is a general enhancement or depression due to the occurrence or absence of pulsating or flaming auroral forms. Similarly it may be seen that such auroral forms are not most frequent when complete absorption is greatest. We may say simply that the occurrence of pulsating and flaming auroral forms or both is often accompanied by complete absorption of vertical incidence waves. There is a closer resemblance between the nocturnal variations of intensity of the aurora and frequency of occurrence of complete absorption than there is between frequency of occurrence of pulsating and flaming auroras and complete absorption.

It is impossible to assess the effect of cloud on the results. Macquarie Island is characterized by extreme cloudiness and rapid changes in cloud cover, and even on nights that are not completely overcast the overwhelming majority has a period when the sky is at least 7/8 clouded for some period. Of 244 nights between August 14, 1950 and April 15, 1951, 205 nights had a period of at least 7/8 cloud between 0000 L.M.T. and 0200 L.M.T.

The author is not able with confidence to join Heppner, Byrne, and Belon in saying that "absorption begins with the appearance of pulsating aurora and ends whenever it is replaced by non-pulsating aurora", because no clear-cut minute to minute association of auroral forms and absorption or abnormal E occurrence has been found. Cloud interference considerably reduced the number of observations and while the remaining observations give some support to the above thesis there is also evidence to the contrary.

IV. ACKNOWLEDGMENTS

This work is part of the research programme of the Australian National Antarctic Research Expedition and could not have been carried out without the observations of A.N.A.R.E. physicists, D. S. Cohen, N. R. Parsons, and K. B. Fenton. The data for College was published by the Central Radio Propagation Laboratory of the U.S. National Bureau of Standards and was made available to the author by the Ionospheric Prediction Service of the Commonwealth Observatory.

V. REFERENCES

- COHEN, D. S. (1952).—Hourly measurements of ionospheric characteristics, Macquarie Island 1950. Aust. Nat. Antarctic Res. Expedition Interim Rep. No. 2.
- HEPPNER, J. P., BYRNE, E. C., and BELON, A. E. (1952).—Association of absorption and E_s ionization with aurora. *J. Geophys. Res.* **57**: 121-34.
- JACKA, F. (1954).—Variations of intensity of the aurora at Macquarie Island. *Aust. J. Phys.* **7**: 477.
- JEFFREY, Z. R. (1953).—Hourly measurements of ionospheric characteristics, Macquarie Island 1951. Aust. Nat. Antarctic Res. Expedition Interim Rep. No. 6.
- PARSONS, N. R., and FENTON, K. B. (1953).—Observations of the Aurora Australis, Macquarie Island, May 1950-Apr. 1951. Aust. Nat. Antarctic Res. Expedition Interim Rep. No. 5.

VARIATIONS OF INTENSITY OF THE AURORA AT MACQUARIE ISLAND

By F. JACKA*

[Manuscript received March 2, 1954]

Summary

Observations of intensity of the aurora at Macquarie Island (geomagnetic coordinates 61°S. , 243°E.) are examined. The intensity is found to be dependent on geomagnetic planetary disturbance index K_p , but the form of this dependence is different for different auroral forms. After eliminating the effects of variation of K_p the residual diurnal and annual variations of intensity are small and irregular.

I. INTRODUCTION

The nocturnal variation of auroral "activity" has been examined by a number of workers (cf. Vegard 1912 ; Fuller 1933 ; Davies 1935) but agreement among them is very poor. The annual variation of auroral activity has been shown to exhibit maxima at the equinoxes and minima at the solstices (cf. Chapman and Bartels 1940, p. 474). A high correlation between auroral activity and magnetic activity has been claimed by some workers and denied by others (cf. Rooney 1934, p. 109). In these studies various measures of auroral activity have been used—in some cases the intensity, in some the frequency of occurrence, and in others a measure depending on intensity, area of sky covered, and type of aurora.

In an earlier paper (Jacka 1953) it was pointed out that the geomagnetic planetary disturbance index K_p was devised to measure the "intensity" of the solar particle stream producing the aurora and it will be inferred from the discussion of that paper that the definition of the term "intensity" must depend on the formulation of a satisfactory theory of the aurora and magnetic disturbance. It is considered that a study of the variations of intensity of the aurora and its dependence on K_p may contribute to this end.

For this purpose Parsons and Fenton's (1953) observations, which cover the period May 1950 to April 1951, have been examined. These observations were made at the Australian National Antarctic Research Expedition station at Macquarie Island (geomagnetic coordinates 61°S. , 243°E.). The intensity I of the brightest display observed during each hour of G.M.T. is considered. These intensities were visually estimated on the usual 0-4 scale (cf. La Cour 1932); consequently they are very roughly proportional to the logarithms of the photometric intensities.

* Antarctic Division, Department of External Affairs, Melbourne.

II. NOCTURNAL VARIATION OF AURORAL INTENSITY

The mean nocturnal variation of I is shown in Figure 1 with the mean nocturnal variation in K_p derived from data associated with the same hours of observation. The values of K_p are taken to the nearest integer and are ascribed to each hour of the standard 3-hr periods. Mean magnetic midnight (as defined

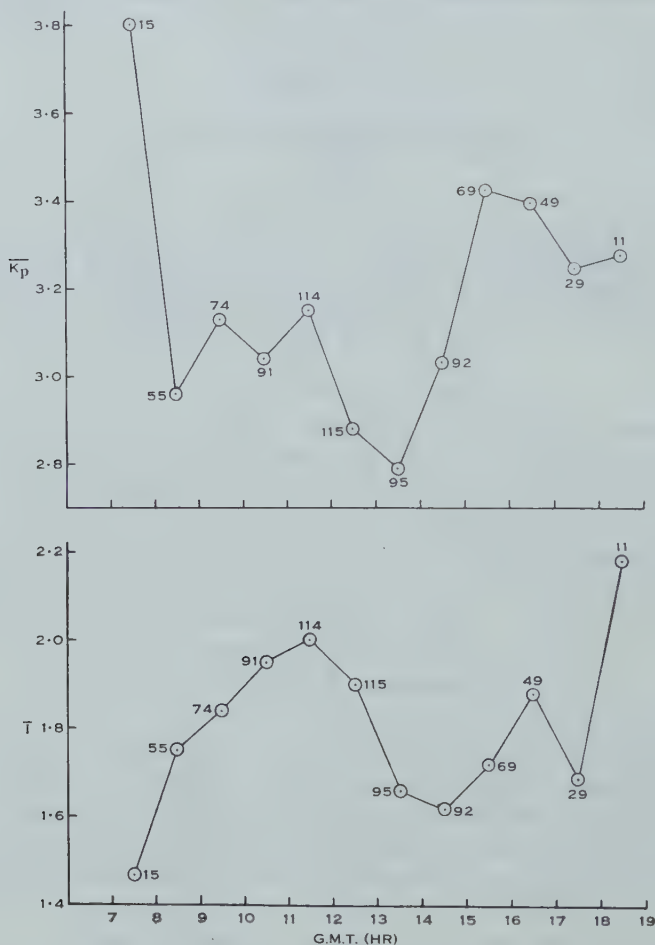


Fig. 1.—Nocturnal variation of intensity of the aurora, I , and K_p .
(The numbers near the points indicate the number of observations used in computing the point.)

by McNish (1936)) is at 12.4 hr G.M.T. and geographic midnight at 13.4 hr G.M.T.

Figure 2 shows the mean value of I associated with each value of K_p , the straight line

$$I_{E1} = 0.57 + 0.39K_p \quad \dots\dots\dots (1)$$

through the points being fitted to the raw data by the least squares method.

Ordering the observations chronologically, we find

$$d = \Sigma(\Delta Z)^2 / \Sigma Z^2 = 1.27 < d_{L(5\%)} \simeq 2,$$

indicating serial correlation of the error term of the regression model significant at the 5 per cent. level (cf. Durbin and Watson 1951). (Z is deviation from regression, ΔZ 's are first differences of Z .)

The mean $I - I_{E1}$ for each hour is shown in Figure 3. The distribution of points here suggests a relation of the form

$$I_{E2} = \alpha + \beta_2 K_p + \beta_3 \sin \frac{2\pi(t-9)}{8}, \quad \dots \dots \dots (2)$$

t being the mid point of the hour (G.M.T.) of observation. The least squares estimates and 95 per cent. confidence limits of the coefficients are $\alpha^* = 0.56$,

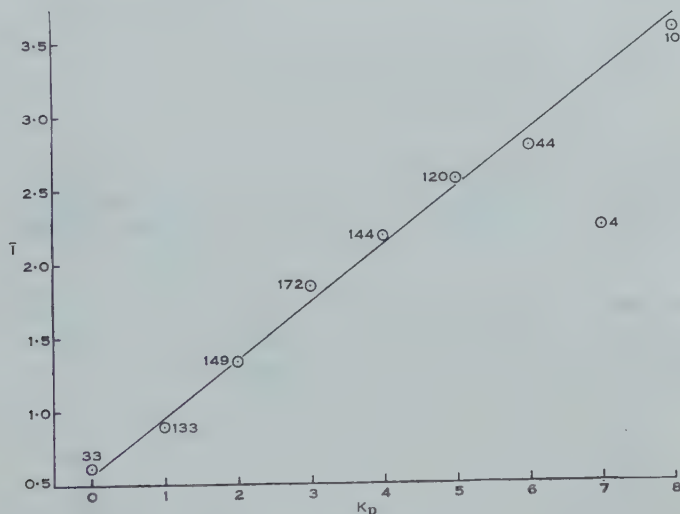


Fig. 2.—Dependence of intensity of the aurora, I , on K_p . (The numbers near the points indicate the number of observations used in computing the point.)

$\beta_2^* = 0.39 \pm 0.04$, $\beta_3^* = 0.14 \pm 0.09$. Again, however, $d = 1.36 < d_{L(5\%)} \simeq 2$, indicating serial correlation of the error term. The significance of the estimates above is, therefore, doubtful especially in the case of β_3^* (cf. Durbin and Watson 1950). The curve $0.14 \sin \{2\pi(t-9)/8\}$ is shown in Figure 3.

From the physical point of view the regression model is unsatisfactory if the error term is serially correlated. In attempting to improve the regression model the following possibilities should be considered :

- (1) The intensity I may be dependent on some factor, other than K_p and t , which is not highly correlated with K_p or t .
- (2) The observers' estimates of I may not be consistent; they may tend to be high for a while—say a few hours—then low for a while and so on.

(3) It may be that the relation between I , K_p , and t is different for different auroral forms and that the form of the brightest display during each hour shows

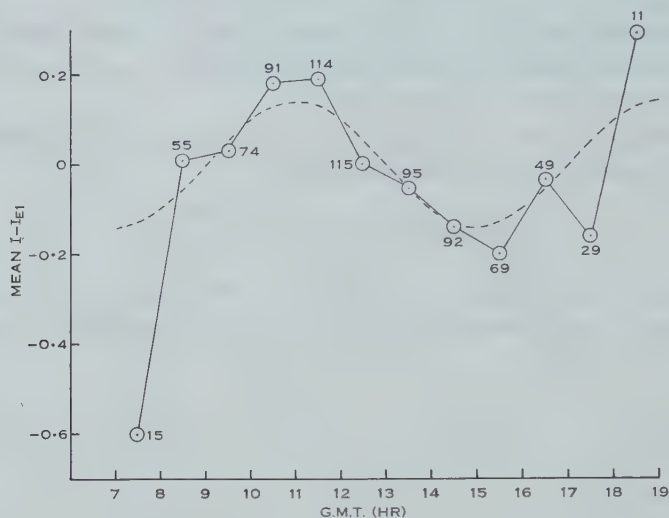


Fig. 3.—Nocturnal variation of deviations $I - I_{E1}$ of observed intensity from regression of I on K_p . The dotted curve is that of $0.14 \sin \{2\pi(t-9)/8\}$. (The numbers near the points indicate the number of observations used in computing the point.)

some tendency to conservation. In fact, an examination of the observers' log (Parsons and Fenton 1953) shows quite clearly that a conservation of the form of the brightest display for several hours is quite common.

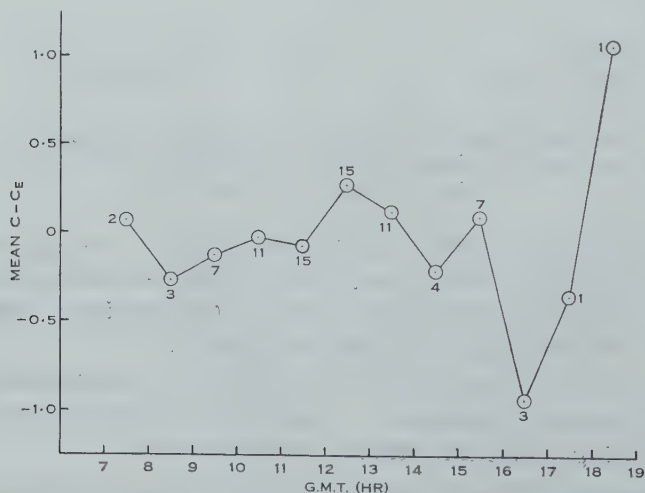


Fig. 4.—Nocturnal variation of deviation $C - C_E$ of observed corona intensities from regression on K_p . (The numbers near the points indicate the number of observations used in computing the point.)

In order to examine the plausibility of this last point the observations on "coronas" and on "homogeneous arcs" (H.A.) were examined. Denoting by C the intensity of the brightest corona during each hour in which a corona was observed we find

$$C_E = 1.18 + (0.29 \pm 0.17)K_p, \quad \dots\dots\dots (3)$$

the limits being the 95 per cent. confidence limits of the coefficient of K_p . Serial correlation of the error term is not significant at the 5 per cent. level;

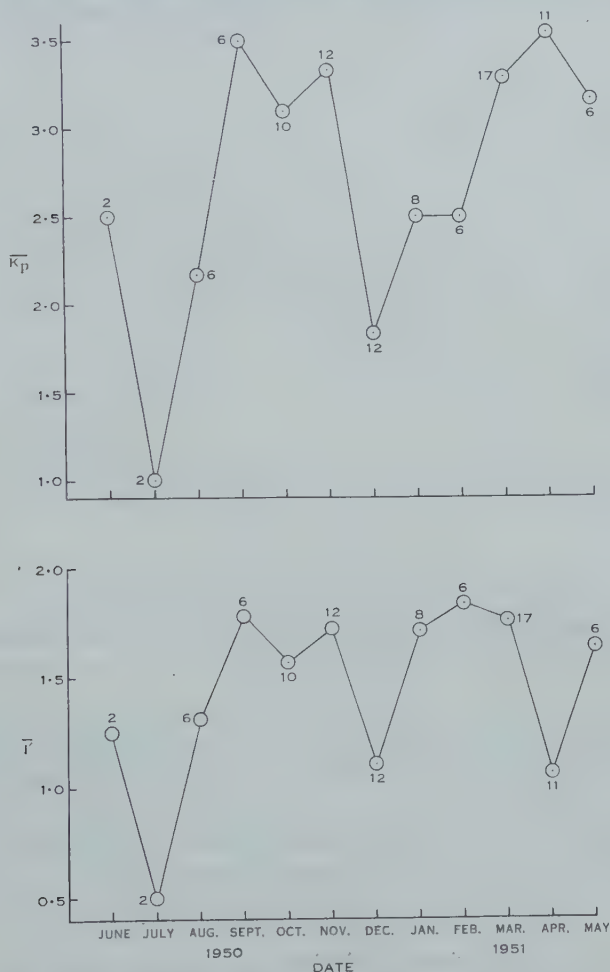


Fig. 5.—Annual variation of monthly mean auroral intensity I' and K_p . (The numbers near the points indicate the number of observations used in computing the point.)

$d = 1.73 > d_{U(5\%)} = 1.66$. (This is not surprising in view of the fact that the observations, 80 in all, were made over a period of 1 year so that there were rarely two or more observations during one night). The variation (which may not be significant) of $C - C_E$ with time of night is shown in Figure 4; its form is markedly different from that of $I - I_{E1}$ shown in Figure 3.

Denoting by A' the mean intensity of all H.A. observed during the one hour we find the regression of A' on K_p and also of A' on K_p and L' (mean latitude of H.A.) is not significant at the 5 per cent. level.

Point (3) above, then, does offer at least a partial explanation of the failure of the regression models considered above. In order to describe the variations of intensity of the aurora we must consider the dependence of intensity on K_p and t for each auroral form and also, separately, describe the sequence and concurrence of forms characteristic of the place of observation. That the

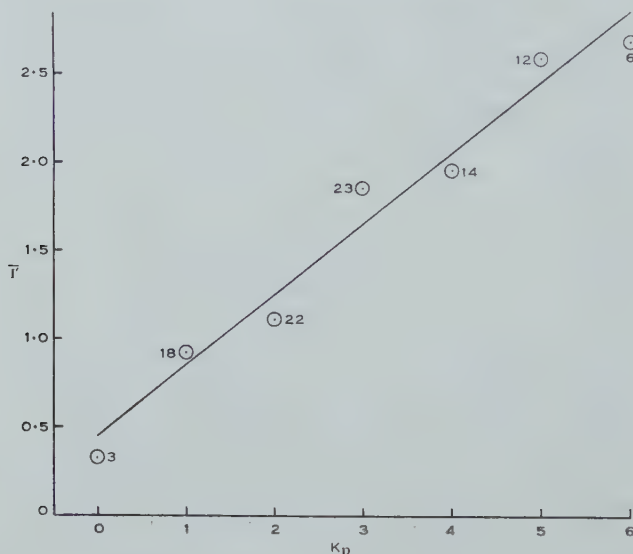


Fig. 6.—Dependence of intensity of aurora, I' , on K_p . (The numbers near the points indicate the number of observations used in computing the point.)

sequence of forms is dependent on the place of observation is clear from the fact that the probability distribution of geographic position is different for different forms (cf. Jacka 1953, p. 227).

It is considered that these findings may explain the poor agreement among previous workers on the diurnal variation of auroral activity.

III. ANNUAL VARIATION OF AURORAL INTENSITY

In order to study the annual variation of intensity of the aurora the mean values I' of the values of I for the hours 12–13, 13–14, and 14–15 hr G.M.T. were calculated for each day. In cases where only two values of I were available in this 3-hr period I' was taken as the mean of these; in cases where only one value was available this was not included in the sample.

Figure 5 shows the annual variation in monthly mean of I' and also of K_p for corresponding 3-hr periods. Figure 6 shows the mean I' associated with each value of K_p , the straight line

$$I_E = 0.45 + (0.40 \pm 0.10)K_p \quad \dots \dots \dots (4)$$

through the points being fitted to the raw data by the least squares method. The limits are the 95 per cent. confidence limits of the coefficient of K_p . In this case serial correlation of the error term is not significant at the 5 per cent. level; $d=1.90 > d_{U(5\%)}=1.69$.

The monthly mean values of $I' - I'_E$ are plotted in Figure 7; this shows the annual variation of intensity of the aurora after elimination of the effects of variation of K_p . The variation is small and irregular and may well be due to errors in estimating the intensity.

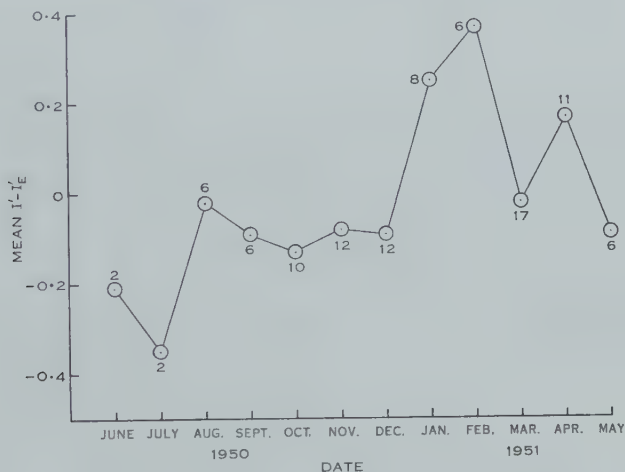


Fig. 7.—Annual variation of deviations $I' - I'_E$ of observed intensity from regression on K_p . (The numbers near the points indicate the number of observations used in computing the point.)

IV. CONCLUSIONS

Even though visual estimates of the intensity of the aurora are very rough and subjective it appears likely that an application of the methods used above on such estimates may lead to descriptions of the variations of intensity of different auroral forms which would be of use in formulating a satisfactory physical theory of the phenomenon. The possibility should not be overlooked, of course, that the magnetic disturbance index K at the place of observation may provide more relevant information on the auroral intensity than does K_p .

V. ACKNOWLEDGMENTS

The author wishes to express his thanks to Mrs. U. Brent and Miss J. Gregory for their assistance with numerical computations and preparation of diagrams and to Mr. N. R. Parsons and Dr. K. B. Fenton for discussions on their observations.

VI. REFERENCES

- CHAPMAN, S., and BARTELS, J. (1940).—“Geomagnetism.” (Oxford Univ. Press.)
 DAVIES, F. T. (1935).—*Terr. Magn. Atmos. Elect.* **40**: 173.
 DURBIN, J., and WATSON, G. S. (1950).—*Biometrika* **37**: 409.

- DURBIN, J., and WATSON, G. S. (1951).—*Biometrika* **38**: 157.
- FULLER, V. B. (1933).—*Terr. Magn. Atmos. Elect.* **38**: 207.
- JACKA, F. (1953).—*Aust. J. Phys.* **6**: 219.
- LA COUR, D. (1932).—"Supplement to the Photographic Atlas of Auroral Forms." p. 12.
(Int. Un. Geod.: Oslo.)
- MENISH, A. G. (1936).—*Terr. Magn. Atmos. Elect.* **41**: 37.
- PARSONS, N. R., and FENTON, K. B. (1953).—Observations of the Aurora Australis, Macquarie Island, May 1950-April 1951. Aust. Nat. Antarctic Res. Expedition Interim Rep. No. 5.
- ROONEY, W. J. (1934).—*Terr. Magn. Atmos. Elect.* **39**: 103.
- VEGARD, L. (1912).—*Phil. Mag.* **23**: 211.

PROGRAMME DESIGN FOR THE C.S.I.R.O. MARK I COMPUTER

III. ADAPTATION OF ROUTINES FOR ELABORATE ARITHMETICAL OPERATIONS

By T. PEARCEY* and G. W. HILL*

[*Manuscript received November 12, 1953*]

Summary

Certain routines, using the fixed index programming methods established for use in the C.S.I.R.O. Mark I computer, are described. These facilitate the use of the machine for performing elaborate arithmetical operations required for extended accuracy, floating index and complex variable arithmetic. Programming for such operations is greatly simplified by the use of an "interpretive" code, especially chosen by the programmer. The "interpretive" method of programme design is potentially very powerful, and relieves the programmer of the work associated with the details of store positioning, index control, scale changes, etc. The various operations called into use by the specially chosen codes are closely analogous to the operations existing in the normal machine code, although some are specially chosen to facilitate the use of variable commands and in transfers of control by linkage and in control of repetitions of routines which are stored in the machine in the special code.

I. INTRODUCTION

The C.S.I.R.O. Mark I computer is of the high speed fully automatic type which is organized and sequenced by a suitable programme of commands. The machine is capable of performing a number of elementary functions, multiplication with fixed index point, and certain logical functions. Each command corresponds to one of these elementary functions. Normally programmes use the system of library routines and sub-routines, which themselves consist of sequences of commands for organizing such operations as division, square rooting, evaluation of functions, and so on.

In Parts I and II (Pearcey and Hill 1953*a*, 1953*b*) the authors described the conventions used in programme design and showed how a programme is compiled and recorded in a manner suitable for use by the computer. The discussion was restricted to cases of single word numbers, i.e. numbers occupying only one storage location, and to cases of fixed index point.

This part discusses the extension of the library routine system to more elaborate arithmetical systems; to floating point, multiple precision, and complex arithmetic, and certain combinations of these. The use of routines providing arithmetical operations of these kinds is greatly facilitated by use of a very flexible method of programme organization termed "interpretive" (Wilkes, Wheeler, and Gill 1951), and a full description of the application of this method to the C.S.I.R.O. Mark I computer will be given.

* Division of Radiophysics, C.S.I.R.O., University Grounds, Sydney.

The overall effect of use of interpretive methods for performing elaborate arithmetical operations is to provide the machine with a number of additional useful functions each of which may be called into use individually by single special commands. The code system used for recording commands in such programmes differs from that used for normal commands used by the machine, and can be chosen at will to suit the system of routines in use.

II. ARITHMETICAL SYSTEMS

Most automatic computers like the C.S.I.R.O. Mark I operate in the fixed index convention, that is, the location of the binary point is fixed in relation to the digits of numbers which are restricted to a definite range of values, usually between $+1$ and -1 . In such a scheme it is often difficult to keep the range of magnitude of variables within the storage capacity of registers. Variables may require to be scaled down initially, thus losing significant digits. Alternatively, suitable changes in the position of the index point may be "programmed"; that is, routines are inserted where needed which cause variables to be scaled up or down during the course of the calculation. Both these expedients are inconvenient and call for considerable effort from the programmer.

Additional difficulty may arise in programming for calculations involving large sequences of arithmetical operations and those involving complex numbers. In the latter case the number of problem variables is at least doubled and organization of the calculation correspondingly increased.

In some types of calculation it is impracticable to predict the range of variables at any particular stage of the calculation. This occurs in evaluation of large determinants, transformation of matrices, and evaluation of roots of algebraic, logarithmic, and transcendental functions. In such calculations a semi-logarithmic form of recording data is most useful. Each datum is recorded in two parts; a fractional component, X say, and an integral component, p . Two scale conventions are commonly in use, the decimal and binary index scales. In the first of these the components are related by a convention which represents the number $X \cdot 10^p$ ($0.1 \leq |X| < 1$; $-N \leq p < N$).^{*} In automatic computers decimal index scale is commonly used but binary index scale is sometimes more convenient when the greatest precision from a given number of binary digits is required. In the latter case, the number represented would be $X \cdot 2^p$ ($0.5 \leq |X| < 1$; $-M \leq p < M$).^{*}

The semi-logarithmic or "floating index" method of storage is quite suitable when the number of multiplications, divisions, square rooting, etc. is comparable with, or greater than, the number of additions and subtractions. The method fails in subtraction of nearly equal quantities, as also occurs in fixed index scales. Significant figures are lost, and can be retained only by holding additional figures in the fractional components throughout the calculation. Conditions can arise in which the result is dependent upon the order in which operations are performed. Such conditions occur in matrix operations, particularly with

^{*} N and M are some integers determined by the maximum capacity of the register storing p .

sets of ill-conditioned coefficients. The floating index method fails also in summation of a large number of items of the same order of magnitude. In numerical integration a stage in the summation may occur at which the partial sum is so great that following increments lie outside the precision of the fractional component, and further increments will not affect the total. The alternative is to perform the addition by a fixed index method and allow the accumulator to "over-carry", account being taken of the digits carried over. This amounts to retaining additional figures.

Floating index arithmetic is therefore not a cure-all for arithmetical difficulties, but must be used with care. Where suitable, it can save much work in programme design and avoids nearly all preliminary scaling adjustments. In case of inadequate accuracy, additional figures must be retained. This can be done in a computer only by allowing the digits of a number to extend beyond one storage location into one or more other locations. Such methods are known as "multiple precision" methods; the commonest being the "double precision" method, in which just two storage locations are used to hold each number. In many cases double precision is sufficient to overcome most of the difficulties mentioned and also to reduce considerably the effort required in programming and in scale adjustments.

In certain cases both the multiple precision and floating index methods may be used together: part of the calculation being carried out by one method and part by the other, or even by using the floating index method together with double or extended precision for fractional components.

These methods may be extended to deal with complex variables. The additional complication is that of organizing the relationships between the two parts of each variable. Arithmetical systems which it may be necessary to use may be listed as follows.

(i) *Fixed Index Methods*.—These may be of single or multiple precision, for real or complex variables or both. Some scale change and progressive adjustments during calculation sometimes cannot be avoided, and changes of index must be specially programmed.

(ii) *Floating Index Methods*.—These may be of single or multiple precision and for real or complex variables or both. No scale changes or index variations are needed nor require to be specially programmed, but the method must be used with care under certain circumstances to avoid loss of significant digits.

III. FUNCTION BLOCKS

For each arithmetical method a set of routines may be constructed for organizing arithmetical and other functions peculiar to the methods adopted. Such a set of routines, known as a "function block", consists of two parts; the "arithmetical block", which provides for addition, subtraction, multiplication, square rooting, and so on, and the "organizational block", which contains routines for organization of programmes using the function block.

(a) *Arithmetical Function Blocks*

Frequently the basic function of any arithmetical block is that of addition. Around this function all the others may be built. Thus, subtraction uses addition; multiplication, particularly in complex and multiple precision methods, uses addition; division uses both multiplication and addition; and so on. Consequently an arithmetical function block consists of a number of interrelated routines of various orders, linkage being made via registers D_{15} , D_{14} , etc.

A number of function blocks have been constructed and are used in programmes on the C.S.I.R.O. Mark I computer. Each system is designed according to certain conventions which define their manner of use. These are:

(i) *Floating Index Arithmetic*.—In this system numbers are stored in two parts, a fractional part, X say, and an integral part, p . X occupies one store location, i.e. 19 binary digits and one sign digit, while p occupies only 10 digits, including its sign digit, of a second storage location, usually the next higher to that holding X . Negative numbers are represented by the complement of the fractional part. There are two scales used, the decimal and binary.

(1) In the decimal system X is stored in binary form and is restricted to the range $1.0 > |X| \geq 0.1$, whilst p , in p_{11} units, is restricted to the range $512 > p \geq -512$. The number pair X, p thus represents $X \cdot 10^p$.

(2) In the binary system X is stored in binary form and is restricted to the range $1.0 > |X| \geq 0.5$, whilst p , in p_{11} units, is restricted to the range $512 > p \geq -512$. The number pair X, p thus represents $X \cdot 2^p$.

Values of X satisfying the appropriate conditions are said to be "normalized".

In the addition of two numbers, say $X \cdot 10^p$ and $Y \cdot 10^q$, the lower index, say q , is raised to equal the greater, the fractional part being correspondingly reduced to $Y \cdot 10^{q-p}$. The sum $X + Y \cdot 10^{q-p}$ is formed. If this exceeds unity in modulus it is multiplied by 10^{-1} and the index of the sum is changed to $p+1$. The sum is stored as one or other of the pairs

$$\begin{aligned} X + Y \cdot 10^{q-p}, p, \\ (X + Y \cdot 10^{q-p})10^{-1}, p+1. \end{aligned}$$

Multiplication is performed by forming the product XY and the sum $p+q$. The fractional part may lie between 0.1 and 0.01, in the decimal case, and, if so, is multiplied by 10 and the index of the product changed to $p+q-1$. In the case of floating binary indices, multiplications by 10 or 10^{-1} are replaced by multiplications by 2 and 2^{-1} respectively to keep the fractional part of the product between 1 and 0.5 in modulus.

Iterative processes are used for division, square rooting, etc.; the results always being provided in the normalized condition in accordance with conventions of storage.

From a purely arithmetical point of view the floating binary method is the better since changes of the index correspond to multiplication by 2 or 2^{-1} , as

against 10 or 10^{-1} in the decimal method ; the smaller factor assists in maintaining greatest possible accuracy from a given number of digits in the fractional part.

Simpler organization of left and right shifts rather than multiplications by 10 and 10^{-1} leads to a saving of about 10 per cent. of commands in arithmetical routines for floating binary compared with those for floating decimal scales. The binary index scale also provides somewhat faster operation. However, users find visual interpretation easier for floating decimal numbers than for floating binary numbers. It is therefore preferable to have data punched and printed in floating decimal form. Conversion to and from floating decimal form in input and output requires additional routines in the floating binary function block ; which more than compensates for the 10 per cent. saving in the arithmetical routines.

The function block for single word floating-index real-variable arithmetic is relatively simple, involving little interconnection between component routines performing the various operations. Thus the routine for subtraction uses addition, but that for multiplication does not, and the division routine is self-contained. The square root routine uses multiplication. All but division also use a "normalization routine" which converts results to the conventional form for storage.

In calculations which involve many separate groups of data, such as coefficients of matrices which cannot be inspected easily, a special floating decimal function block is used. In this a tally is kept of changes in the number of figures which are significant in the associated number. Thus, if a sum or difference is multiplied by 10^s ($s > 0$) to bring it to the normalized form, the number of significant figures is reduced by s , and so on for other operations. The tally of such changes is held as an integer in p_1 units in the p_1 - p_{10} digit group of the word holding the index. Results can then be given a degree of significance automatically.

(ii) *Multiple Precision Arithmetic.*—In this system a number is divided into groups of 19 binary digits. The number of such groups may be made as great as desired, and a number occupying " n " groups is called an " n -fold" number, and the arithmetical system used with such numbers is called n -fold arithmetic. The most frequently used is the "double precision" case, for which $n=2$. The storage conventions adopted are that an n -fold number is stored in groups of 19 digits in adjacent successive storage locations, the most significant digit position in each location being zero, except in that holding the most significant group, in which it plays the role of a sign digit. A negative n -fold number is stored as a complement of the corresponding n -fold positive number.

Arithmetical blocks are designed in accordance with the convention that n -fold numbers are of unit magnitude or less, the index point being immediately to the right of the most significant p_{20} digit stored.

Two function blocks have been constructed, one for double precision and one for n -fold arithmetic, where n may be specified by the programmer. The latter system finds application in number theory computations.

(1) *Double Precision.* In this method numbers are stored modulo 2 in integral multiples of 2^{-38} , providing an equivalent accuracy of about 12 decimal digits. When addition is performed, digit groups of corresponding significance are added together starting with those of lower significance. Any over-carry arising from summation of the less significant components is detected and added as a unit to the sum of the more significant components. The sum is finally stored in the standard form.

The function block provides the functions of addition, subtraction, multiplication, division, and square rooting. Each function routine uses the routine for addition, and the last two use also multiplication, which itself uses addition. A product is always provided in the standard form as a group of *four* 19-digit words, the two least significant words being retained in case of need in the calculation. This allows double precision numbers to be treated, by suitable programming, as integers instead of fractions.

The organization of this arithmetical block is more complicated than that for the floating index method since there is a larger proportion of high order routines.

(2) *n-Fold Arithmetic.* In this case n may be specified by the user, but only the functions of addition, subtraction, and multiplication are provided. The organization is correspondingly more complicated than that for the double precision case. Its manner of use differs from that of other systems and will be discussed later.

(iii) *Complex Variable Arithmetic.*—The complications caused in the construction of programmes to deal with the arithmetic of the complex variable justify the use of a function block even in the case of single-fold or one-word arithmetic. This has been done for use on the Mark I computer in a special case which was found to be most frequently used. This case adopts the convention that the real and imaginary parts of a complex number are stored in two adjacent storage locations. The index points are considered fixed and to lie between the p_{10} and p_{11} digit positions in both components and *not* between the p_{19} and p_{20} digit positions. Negative components are represented by the complement of the corresponding positive number. Each component of a number may take values equal to all integral multiples of 2^{-10} from -512 to $512 - 2^{-10}$ and all the routines of the function block preserve this convention.

Routines contained in the function block include addition, subtraction, multiplication, division, and square rooting of complex numbers, together with other functions which apply only in the case of the complex variable, such as the evaluation of the modulus and the conjugate of a complex number.

Function blocks are also available for use of the complex variable in :

- (1) floating index arithmetic,
- (2) double precision arithmetic.

In each case the same conventions apply to both components as apply also to the corresponding case of the real variable. One exception is the floating

index system where only three adjacent storage locations are used instead of four. The first and third hold the real and imaginary fractions, respectively, the second or centre location holds the indices of both components, that of the real in the p_{11} - p_{20} positions and that of the imaginary in the p_1 - p_{10} positions. Whenever such a number is taken from store the indices are "unpacked" before the arithmetical operation is performed. When placed into store the indices are "repacked".

In all cases routines for the complex variable are based upon routines in the corresponding function block for the real variable. Thus complex floating addition uses the routine for real floating addition, complex floating multiplication uses real floating multiplication and real addition, and so on, and all routines for complex operations are of higher order than the routines for real operations.

Further, since it frequently happens in calculations involving the complex variable that operations with real variables are also needed, the "real" operations are also provided in addition to the additional functions of modulus and conjugate, etc. which are peculiar to the complex variable.

The principles of the construction of a block for n -fold complex arithmetic would be similar to those already constructed.

(iv) *Other Systems.*—The most likely useful system in addition to those described above is the floating double precision system and its counterpart for use with the complex variable. In this the conventions of both the floating index and double precision methods are combined. Each number occupies two adjacent storage locations for its fractional part and its index lies in the p_{11} - p_{20} positions in the following location. A complex number occupies five locations, the first two for the real and the last two for the imaginary fractions, the indices being packed into the third location. The function blocks for these systems are built in direct analogy to those for the other systems. Both function blocks are necessarily large and occupy much storage space, and also are slower in operation since in effect all the operations of both the floating index and double precision systems must be performed.

(b) *Organizational Function Blocks*

Into the organizational part of the various function blocks are placed routines for taking fresh data from punched tape, page printing, or punching results on to tape; and for withdrawing numbers from and placing numbers into store in standard form appropriate to the system used. For purposes of organizing the problem programme certain other operations are also included. These are analogous to the control shifts and counts which occur in normal routines and will be illustrated in more detail later. Of these operations only the input and output functions use the arithmetical block. Input tape for floating index arithmetic is punched with the fraction preceding the index. A designation follows each component to denote its sign. Decimal digits are received from tape and converted by the input routine of the function block and assembled into the equivalent binary form and the result is stored. For

printing, the fractional part of a number is printed as six decimal digits with sign followed by the index printed as a three-digit decimal integer with sign. The input or output operations read or print data in standard form appropriate to the function block used. In the case of the complex variable, the real part always precedes the imaginary part.

The manner of placing into and taking from store is standardized according to the system's conventions. Thus, routines take and place the various parts of a number from and into successive storage locations in the correct standard order, doing the packing or unpacking of parts as required.

IV. USE OF FUNCTION BLOCKS

The most straightforward way of using a function block is to call the various function routines into use by the standard link and cue method described in Part II (Pearcey and Hill 1953*b*). A problem programme would then consist largely of "control commands", those which affect the sequence register and cause control to be stored or shifted, together with relatively few commands for operations which may not be provided by the function block and by other routines included into the programme.

This method suffers from the disadvantage that the link number must frequently be restored as routines of fresh orders are called into use, and that the withdrawal and placing into store of numbers in standard form must be made with reference to an address which must, each time, be previously placed into a special location. These factors are confusing for the programmer; they make it more difficult to trace through the sequence of operations in a programme and draw his attention away from the fact that the function block really supplies the machine with what in effect is a new set of "wired-in" operations.

One way of helping the programmer is to provide a "directory" with the functions blocks. This consists essentially of a list of numbered addresses placed in sequence in the store; each address corresponds to the place in the function block, to which control must be transferred in order to obtain the operation which the particular directory number corresponds to.

The use of directory numbers for calling functions in the block implies the use of special operation codes, not identical with the machine code.

This development leads directly to the "interpretive system" of programming, a method which greatly reduces effort needed to compile programmes.

V. THE INTERPRETIVE SYSTEM

Each function called via the directory may be coded into a programme in terms of the serial number of the directory position which causes control to be transferred to the required function routine, and a special routine is needed for interpreting those parts of a programme which are recorded in the special code.

The particular code adopted for interpretation is at the choice of the programmer; and is not necessarily restricted to any particular address system.

Each coded word or partitioned group of digits may be regarded as a command of a special kind, the meaning of which may change from problem to problem and is not directly connected with the actual mode of operation of the machine. The special decoding or "interpretation routine" may be designed at the will of the programmer in accordance with his chosen code system.

The simplest way of using a list of coded data in a programme is to arrange the interpretation routine so as to select and interpret them one by one in order of storage location. The list of coded data may thus be considered as a routine in its own right. Many of the techniques used in the design of normal routines may be taken over into the design of routines to be interpreted and a strong parallelism can be drawn between normal techniques and interpretation techniques. Terms which are used in normal programming are frequently adopted with similar meaning in the interpretive method and are prefixed by the term "hyper-". Thus the list of programme data for interpretation is known as a hyper-routine or hyper-programme and the code adopted is called the hyper-code; the function corresponding to any hyper-code number is a hyper-function. A single datum of a hyper-programme is called a hyper-command. If the datum to be currently interpreted is stored in location n , then hyper-control is said to be at n .

Similarly, some of the registers existing in the computer, particularly the accumulator A , have their equivalent in the interpretive system. These are known as hyper-registers, for example, hyper-accumulator, hyper-sequence register, hyper-interpreter, etc. They must not be confused with the accumulator, sequence register, or interpreter proper. Thus the physical location of the hyper-accumulator depends upon the function block used, and may in fact be one or more of the registers of D or some locations of the store.

VI. THE HYPER-CODE

Many hyper-code systems are possible and the programmer may design whichever suits his purpose. In most interpretive systems it is convenient to code a single hyper-command into a single location of the store, although if additional digits are needed two or more adjacent locations in the store may be used to hold a single hyper-command; or more than one hyper-command, if requiring few digits each, may be packed two or more together into single locations. The digits of each hyper-command are partitioned into groups each possessing a meaning depending upon the structure of the interpretation routine. The parallelism between normal, or routine, commands and hyper-commands is extended to the address conventions of the hyper-code, which may be of the one-address or three-address type and so on. A one-address hyper-code is that most commonly used with the C.S.I.R.O. Mark I computer.

A typical one-address hyper-command of one word consists essentially of a numerical address of 10 digits occupying the digit positions p_{11} - p_{20} and a function address of five digits in the positions p_6 - p_{10} . The digit group p_1 - p_5 in hyper-commands corresponds to the machine code for destination Z , or number 20, a destination which in the machine code has been left unused. Any machine

command possessing this destination has no effect except in the case of the $c(A)$ source, when register A becomes cleared to zero.

The destination Z , or 20, distinguishes those hyper-commands which call hyper-functions from the function block from those which do not. It is not always convenient to construct hyper-routines using only the operations available in the function block; occasionally quite simple operations are required for which it is not economical to provide store space for a short routine in the function block. Hyper-commands not terminated with the Z code letter are coded as machine commands and the interpretation routine deals with them as though they were machine commands. In this way, what are in effect machine commands may be entered into hyper-routines. Such commands are true hyper-commands but do not call the function block into use.

The numerical address of a Z -terminated hyper-command in the p_{11} - p_{20} digit group, possesses the same significance as it does in a machine command. The function address in Z -terminated hyper-commands is a serial number, from 0 to 31, providing space for 32 possible functions, some of which are essentially arithmetical and others organizational. This code number is selected from the hyper-command by the interpretation routine and used as a reference number to the directory, which is contained in the interpretation routine and which in turn directs control to the required position in the function block.

VII. p -WORD NUMBERS

In the conventions of many function blocks a datum occupies more than one word of the store. Data may occupy different numbers of locations in the store for the same function block. Thus for complex arithmetic complex numbers require more words than real numbers in the same problem. For convenience the multi-word locations $n, n+1, n+2, \dots, n+p-1$ are referred to as " p -length location n " and symbolized as n_p .

For programmes using real and complex variables, two word lengths will occur; p -length words for real numbers and q -length words for complex numbers. However, by incorporation of special features into the function blocks the programmer is relieved of any detailed concern with such "mixed" word lengths.

VIII. HYPER-REGISTERS

By analogy with the registers used in normal programming the hyper-registers serve equivalent purposes in hyper-programmes. The hyper-address code is of the one-address type, there is only one arithmetical hyper-register known as the hyper-accumulator. It is denoted by \bar{A} . The hyper-sequence register holds the store address of the next hyper-command to be adopted and is denoted by \bar{S} , and the hyper-interpreter, denoted by \bar{K} , holds the numerical address of the current hyper-command. A further hyper-register known as the "transfer register", denoted by X , is used for transfers of numbers between store and the accumulator. A further hyper-register, which, like X , has no parallel in normal programming, is the hyper-link register, written \bar{L} . The

physical position of these registers varies with the function block used. As an example, those for the floating decimal index system are listed below :

Hyper-accumulator	$\bar{A} = D_{10}D_{11}$ for fraction and index respectively
Transfer register	$X = A, B$ for fraction and index respectively
Hyper-sequence register	$\bar{S} = D_{13}$
Hyper-interpreter	$\bar{K} = D_{12}$
Hyper-link register	\bar{L} = three spaces in the function block
Normal link-stores	$= D_{14}, D_{15}$, used for links into and between function block routines

Stores for constant parameters in function blocks $= D_8, D_9$, etc.

Stores for count parameters in hyper-programmes $= D_0, D_1, \dots, D_7$

The advantage gained by placing hyper-registers into D whenever possible is the ease with which the operator may see, at a glance at the monitor tube faces, at which stage the machine is in the hyper-programme. With the more complicated function blocks it is necessary to extend the space occupied by hyper-registers into the function block itself.

IX. HYPER-CONTROL AND HYPER-ROUTINE LINKAGES

Most of the techniques used in normal programme design are applied also to the design of hyper-programmes. However, some special functions are provided in the function blocks which simplify somewhat the design of hyper-programmes, and have no direct equivalent in normal programming methods.

Such is the case of those hyper-commands devoted to control of the hyper-sequence and linkage to and from hyper-routines. The hyper-registers involved in sequence control are known as : the hyper-interpreter \bar{K} , the hyper-sequence register \bar{S} , and the hyper-link register \bar{L} . The last of these comprises a group of consecutive storage locations and may be called $\bar{L}_1, \bar{L}_2, \bar{L}_3$, etc.

When control enters the interpretation routine hyper-control advances serially from the address held in \bar{S} at the stage of entry. During each cycle of the interpretation routine a p_{11} unit is added to \bar{S} , e.g. to D_{13} .

A change of hyper-control to " n " is called by a hyper-command written as

$$n \rightarrow \bar{S}$$

coded as

$$n; m, Z,$$

where m is the address in the directory which calls the appropriate function in the function block. This function places the content of \bar{K} (or D_{12}) into \bar{S} (or D_{13}).

In this sense \bar{K} and \bar{S} are respectively analogous to real interpreter and sequence registers. Addition into \bar{S} is not provided.

Transfer of hyper-control from one hyper-routine to another and back uses the hyper-link register \bar{L} . The hyper-command $n \rightarrow \bar{L}$ at location m , causes

hyper-control to be transferred to n , and the number $(m+1)p_{11}$ is stored in \bar{L}_1 whilst the previous content of \bar{L}_1 is placed in \bar{L}_2 , and that of \bar{L}_2 goes to \bar{L}_3 , and so on. Such a hyper-command corresponds to the sequence

$$(S) \rightarrow D_{15}, \\ n \rightarrow S$$

in normal programming.

A return of hyper-control, to $m+1$, that following the place from which hyper-control was transferred, is obtained by a hyper-command written as $(\bar{L}) \rightarrow \bar{S}$.

This causes (\bar{L}_1) to be placed in \bar{S} , (\bar{L}_2) to be transferred to \bar{L}_1 , (\bar{L}_3) to go to \bar{L}_2 , and so on. This is analogous to the sequence

$$p_{11} \rightarrow D_{15}, \\ (D_{15}) \rightarrow S$$

in normal programming.

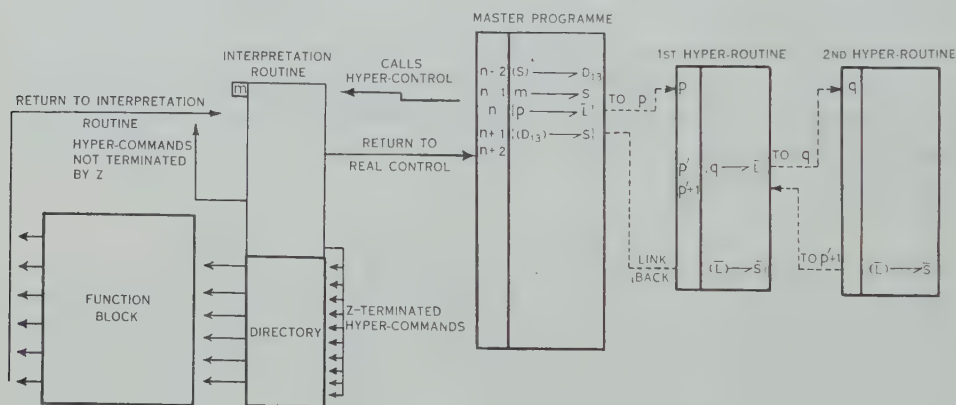


Fig. 1.—Transfers of real control and hyper-control.

Passage of hyper-control ----->

Passage of real control —————>

Hyper-control is called into use by transferring control direct to the interpretation routine in the usual way, that is, by direct transfer of real control to the head of the interpretation routine. To cease hyper-control and achieve return of real control to the main programme a link datum must have been stored in D_{13} ; return to real control is attained by a command $(D_{13}) \rightarrow S$ in the hyper-routine; a hyper-command not terminated by Z and therefore possessing the significance of a machine command.

For example, consider the case of transfer from real control at n to hyper-control at p , which in turn calls a second hyper-routine at q . Let the interpretation routine be entered at m . The scheme achieving this is illustrated in Figure 1, where transfers of real control are shown by full lines, and those of hyper-control by broken lines.

In the figure hyper-commands are distinguished from real commands by surrounding all the former by braces thus: $\{ \dots \}$. Hyper-commands not terminated in Z are also distinguished in this manner.

X. HYPER-ROUTINE LOOPS

As in normal programme techniques, so also in hyper-programmes full use is made of the method of controlling repetitions of groups of hyper-commands by a counting and sign-testing process.

Two functions provided in the organizational part of the function block are used for this purpose. The first is written thus

$$n_p \rightarrow D_r,$$

the numerical part of which consists of the partitioned number $8n+r$, where $0 \leq r < 8$ and $0 \leq n < 64$; the second function is represented by

$$n_p \rightarrow D_r,$$

where n and r are coded in the same manner.

The first function places the number $np.p_{11}$ units into the upper half of register D_r , whilst the second subtracts $np.p_{11}$ from (D_r) , tests the difference resulting and, if negative, increases \bar{S} by an additional unit p_{11} , otherwise not.

The first of these hyper-commands is used to set the number of repetitions of the programme loop to which it refers, the latter combines the functions of counting repetitions and causing a conditional shift of hyper-control.

XI. VARIABLE HYPER-COMMANDS

A special function is provided which simplifies the use of variable commands. Such a requirement is frequently associated with the loop control functions just described. This function causes a parameter to be transferred to \bar{K} , and to be held there so that the numerical part of next hyper-command which follows is added into it. This function is denoted by the symbol

$$(D_r) \pm \bar{K}, \text{ where } 0 \leq r < 8.$$

If (D_r) is equal to $np.p_{11}$, the numerical address of the following hyper-command is effectively increased by $np.p_{11}$.

When an operation in a loop of a hyper-routine has its numerical address progressively changed, e.g. an address referring to serial hyper-locations in the store, the address is frequently used also as a counter for loop control. Thus the loop control commands also serve the purpose of assisting in the use of variable commands. In this case it is normally required to vary the address by multiples of p for p -length words and by multiples of q for q -length words in complex operations. Thus two sets of loop control operations are provided for complex variable hyper-programmes

$$\begin{array}{ll} n_p \rightarrow D_r, & n_p \rightarrow D_r, \\ n_q \rightarrow D_r, & n_q \rightarrow D_r, \end{array}$$

having the effects previously described; the second set being used for loops involving addresses for complex operations. For both p and q systems, the same operation $(D_r) \pm K$ can be used where required.

Functions of the type $(D_r) \pm K$ may not be placed adjacent to one another in a hyper-routine.

XII. TRANSFERS TO AND FROM STORE

One hyper-function transfers data to the hyper-accumulator and another transfers from the hyper-accumulator to the store. The former of these is written as

$$(n_p) \rightarrow \bar{A},$$

where n_p is the store location referred to; and similarly the second is written

$$(\bar{A}) \rightarrow n_p.$$

Closely associated with these are the two most elementary arithmetical functions involving transfer from the store to the hyper-accumulator. These are

$$(n_p) \overset{+}{\rightarrow} \bar{A},$$

which adds the content of "hyper-location n " into the hyper-accumulator and

$$(n_p) \overset{-}{\rightarrow} \bar{A},$$

which subtracts the content of hyper-location n into the hyper-accumulator and so on.

In cases of complex arithmetic equivalent q -length operations are supplied for transfers of q -word complex variables etc. Thus the operations provided would include

$$(n_q) \rightarrow \bar{A}; (\bar{A}) \rightarrow n_q; (n_q) \overset{+}{\rightarrow} \bar{A}; (n_q) \overset{-}{\rightarrow} \bar{A}.$$

Table 1 illustrates use of loop control and command variation operations and of braces, $\{ \}$, to distinguish hyper-commands from machine commands. This routine accepts numbers from punched tape and stores them in sequential p -word store locations.

TABLE I
USE OF LOOP CONTROL, COMMAND VARIATION OPERATIONS, AND BRACES

Location	Hyper-command	Operation
0	$\{n-1_p \rightarrow D_r\}$	Sets $(n-1)p.p_{11}$ in D_r
1	$\rightarrow \{(D_r) \overset{+}{\rightarrow} \bar{K}\}$	Adds $(n-s-1)p.p_{11}$ to next command
2	$\{(\bar{I}) \rightarrow m\}$	Stores input number in p -word starting at $m + p(n-s-1)$
3	$\{1_p \overset{-}{\rightarrow} D_r\}$	Reduces (D_r) by $p.p_{11}$ and tests the sign of (D_r)
4	$\rightarrow \{1 \rightarrow \bar{S}\}$	Return hyper-control to 1
5	$\{(\bar{L}) \rightarrow \bar{S}\} \leftarrow$	Link to superior routine when $(D_r) < 0$.

This sequence causes the variable hyper-command at 2 to be adopted as $\{(I) \rightarrow n-s-1.p+m\}$ on the $(s+1)$ th cycle of the loop of n cycles.

XIII. ADDITIONAL HYPER-FUNCTIONS

Additional hyper-functions may be provided simply by placing appropriate routines, additional to the function block, into the store. These are performed by a special hyper-command coded as $n; 15, Z$ which transfers control to the head

location of the additional routine at location n . Since the address digits of such a hyper-command are used in specifying n , additional hyper-functions cannot refer to store addresses of data and are frequently of the type written as

$$\{f(\bar{A}) \rightarrow \bar{A}\},$$

where $f(\bar{A})$ may be $\log(\bar{A})$, $\exp(\bar{A})$, $(\bar{A})^{1/3}$, etc.

Similar hyper-functions are already included in the normal function block, e.g. $(\bar{A}) \rightarrow 0t$, $(I) \rightarrow \bar{A}$, etc.

XIV. THE STANDARD HYPER-CODE

The hyper-code has been standardized so as to apply to most of the function blocks so far designed; the same code calls the same functions whether the floating decimal complex or double length complex system is used. Variations occur in the additional functions called by the hyper-command described in the previous section.

In the present scheme, the p_5 - p_1 digits are used to distinguish between hyper-commands which use the function block and those which do not. For hyper-commands in which this group is not Z , the conventions of normal machine operations apply as outlined in Parts I and II (Pearcey and Hill 1953*a*, 1953*b*).

The earliest 16 of the 32 code numbers in Z -terminated hyper-commands refer to arithmetical functions and transfers of real variables, and organizational hyper-commands, whilst some of the remaining 16 call the hyper-functions relating to the complex variable.

In calculations involving only real variables the latter group of hyper-functions and its corresponding function block are not used. The first part only of the standard code is used. This, and the fact that the code applies equally well to a number of systems, means that the programmer only rarely needs to take into account his arithmetical system whilst programming.

The hyper-code used is shown in Table 2. It includes the hyper-functions for the complex variable. The first six hyper-commands 0-5 provide transfer to \bar{A} and arithmetical functions of the real variable, the appropriate word length suffix being included (here written as p). These functions include addition, subtraction, multiplication (product of (\bar{A}) by (n_p)), and square rooting. The next three, 6-8, provide input, output, and transfer from \bar{A} to store; the following three, 9-11, provide sequence control operations; the next three, 12-14, refer to loop control. Number 15 calls any additional hyper-functions.

The arithmetical functions for the complex variable, called by code numbers 16-21, are similar to those for the real variable called by 0-5. Input and output of the complex variable are provided by 22-24, and functions peculiar to the complex variable and those involving both real and complex variables together are given the highest code numbers. These include such operations as multiplication and division of a complex by a real quantity, evaluation of the modulus and conjugate of a complex number, and so on.

Each function block can provide different groups of possible additional functions without change to the block. These depend upon the nature of the

TABLE 2
THE STANDARD HYPER-CODE

Code	Hyper-command Symbol	Meaning
$n: 0, Z$	$(n_p) \rightarrow \bar{A}$	Transfers hyper-word (n) to $(n+p-1)$ to \bar{A}
$n: 1, Z$	$(n_p) \pm \bar{A}$	Adds content of hyper-word (n) to $(n+p-1)$ to (\bar{A})
$n: 2, Z$	$(n_p) \bar{\rightarrow} \bar{A}$	Subtracts content of hyper-word (n) to $(n+p-1)$ to (\bar{A})
$n: 3, Z$	$(n_p) \times \bar{A}$	Replaces (\bar{A}) by its product with hyper-word (n) to $(n+p-1)$ (fixed index systems hold less significant part of product in register \bar{B})
$n: 4, Z$	$(n_p) \div \bar{A}$	Replaces (\bar{A}) by its quotient with regard to (n) , $(n+1)$, . . . , $(n+p-1)$
$n: 5, Z$	$(n_p)^{\frac{1}{2}} \rightarrow \bar{A}$	Places square root of hyper-word (n) to $(n+p-1)$ in \bar{A}
$n: 6, Z$	$(n_p) \rightarrow \bar{0}$	Print hyper-word in (n) to $(n+p-1)$ in standard form
$n: 7, Z$	$(\bar{I}) \rightarrow n_p$	Read one standard hyper-word from tape and place in (n) to $(n+p-1)$
$n: 8, Z$	$(\bar{A}) \rightarrow n_p$	Transfers (\bar{A}) to hyper-register (n) to $(n+p-1)$ in standard form
$n: 9, Z$	$n \rightarrow \bar{S}$	Transfer hyper-control to n
$n: 10, Z$	$n \rightarrow \bar{L}$	Transfer hyper-control to n and store present hyper-control number plus one in L
$n: 11, Z$	$(\bar{L}) \rightarrow \bar{S}$	Transfer hyper-control to last location stored in L
$n: 12, Z$	$n_p \rightarrow D_r$	Place hyper-count number n_p in D_r
$n: 13, Z$	$n_p \bar{\rightarrow} D_r$	Subtract n_p hyper-count units from D_r
$n: 14, Z$	$(D_r) \pm \bar{K}$	Form the numerical address of the next hyper-command by adding (D_r) to it
$n: 15, Z$	Specified as desired	Transfer control to n but return to interpretation routine
$n: 16, Z$	$(n_q) \rightarrow \bar{A}$	Transfer (n) to $(n+q-1)$ to \bar{A} as a complex number
$n: 17, Z$	$(n_q) \pm \bar{A}$	Add (n) to $(n+q-1)$ to \bar{A} as a complex number
$n: 18, Z$	$(n_q) \bar{\rightarrow} \bar{A}$	Subtract (n) to $(n+q-1)$ from \bar{A} as a complex number
$n: 19, Z$	$(n_q) \times \bar{A}$	Replace (\bar{A}) by the complex product with (n) to $(n+q-1)$
$n: 20, Z$	$(n_q) \div \bar{A}$	Replace (\bar{A}) by the quotient with regard to the complex number (n) to $(n+q-1)$
$n: 21, Z$	$(n_q)^{\frac{1}{2}} \rightarrow \bar{A}$	Replace (\bar{A}) by the square root of the complex number (n) to $(n+q-1)$
$n: 22, Z$	$(n_q) \rightarrow \bar{0}$	Print the complex number in (n) to $(n+q-1)$ in standard form
$n: 23, Z$	$(\bar{I}) \rightarrow n_q$	Read one standard complex number from tape and place into location (n) to $(n+q-1)$
$n: 24, Z$	$(\bar{A}) \rightarrow n_q$	Transfer (\bar{A}) to (n) to $(n+q-1)$ as a standard complex number
$n: 25, Z$	$ (n_q) ^2 \rightarrow \bar{A}_p$	Replace (\bar{A}) by the squared modulus of the complex number (n) to $(n+q-1)$
$n: 26, Z$	$(n_p) \times \bar{A}_q$	Replace the complex number in (\bar{A}) by its product with the real number in (n) to $(n+p-1)$
$n: 27, Z$	$(n_p) \div \bar{A}_q$	Replace the complex number in (\bar{A}) by its quotient with regard to the real number in (n) to $(n+p-1)$
$n: 28, Z$	$n_q \rightarrow D_r$	Place hyper-count number n_q in D_r
$n: 29, Z$	$n_q \bar{\rightarrow} D_r$	Subtract $n_q.p_{11}$ units from D_r .

routines within the blocks and frequently provide additional input and output facilities; operations upon (\bar{A}) and on (X) and so on. These vary with the function block and are not listed in Table 2.

XV. USE OF THE HYPER-CODE

Diverse code systems for hyper-commands and machine commands may seem a disadvantage. It is, however, uneconomical to construct an interpretive routine whose hyper-code corresponds directly to the machine code and, moreover, such a system would involve frequent transfers from hyper-control to machine control and back in order to perform machine commands.

The apparent disadvantage of use of symbols for written hyper-commands differing from the code as it appears on tape is more than outweighed by the convenience of largely mnemonic symbols applied to all hyper-routines in the design stages.

It should be noted that $\{n \rightarrow \bar{S}\}$, which transfers hyper-control to n , is different from $\{n \rightarrow S\}$ or $\{(D_{13}) \rightarrow S\}$ which transfers real control without preparing to link back to the interpretation routine. In fact $(D_{13}) \rightarrow S$ switches back to normal machine operation; the subsequent commands being performed normally, not interpretively. Also, provision is normally made for transforming single counts to S into single counts to \bar{S} . Thus $\{s(A) \xrightarrow{c} S\}$ becomes in effect a function which may be written as $s(A) \xrightarrow{c} \bar{S}$, which causes hyper-control to advance by an additional unit if (A) is negative.

When coding hyper-routines on to tape full use is made of a special "input control routine H " (of 20 commands) which performs all the control operations of control D (listed in Pearcey and Hill 1953b) and two additional operations " P " and " Q ". The punch configurations, $P=31, 3Y$ and $Q=31, 15Y$ are placed immediately after the commands to which they refer. The effect of such control designations is to multiply the written address by p or q respectively, and then include the effect of any A type designation, punched immediately before that hyper-command, before inserting it into store. Thus with mp_{11} stored in the $1A$ location: $1A\{(n) \rightarrow \bar{A}\} P$, punched as $0,1:7Y:n:0,Z:31,3Y$; is stored as $\{(m+np) \rightarrow \bar{A}\}$, and $1A\{(\bar{A}) \rightarrow n\} Q$, punched as $0,1:7Y:n:8,Z:31,15Y$, is stored as $\{(\bar{A}) \rightarrow m+nq\}$. These special control designations, which automatically transform hyper-commands into the required form, together with the p -word and q -word loop control operations, virtually free programmers from concern with word lengths during design of programmes.

Insertion of the function block is usually performed in the order: primary B routine (from stepping switches); control H routine (from tape); spaces left for insertion parameters; first order function routines: second order function routines (which use first order ones); . . .; highest order function routines; sequence and loop control routines; interpretation routine; directory; additional hyper-function routines.

XVI. EXAMPLE OF A HYPER-ROUTINE

As an example of a typical hyper-routine consider the routine required for the evaluation of a polynomial, of degree n , of the complex variable together with its derivative. The coefficients will be assumed to be complex numbers.

The method of computation by recurrence relation is adopted. Thus, if the independent variable be z and the coefficients be denoted by a_r , then p_r and p'_{r+1} are defined by :

$$\begin{aligned} f(z) &= z^n + a_1 z^{n-1} + a_2 z^{n-2} + \dots + a_n, \\ p_0 &= 1, \quad p_r = p_{r-1} z + a_r, \quad (r=0, 1, 2, \dots, n-1) \\ p_n &= f(z), \end{aligned}$$

and

$$\begin{aligned} p'_0 &= 0, \quad p'_{r+1} = p'_r z + p_r, \quad (r=0, 1, 2, \dots, n-1) \\ p'_n &= f'(z), \end{aligned}$$

where $f(z)$ is the function and $f'(z)$ its derivative with regard to z .

TABLE 3

A HYPER-ROUTINE FOR COMPUTING A POLYNOMIAL OF A COMPLEX VARIABLE AND ITS DERIVATIVE

Location	Symbol Code	Action	Tape Code
0 1S 3A,	$\{(\bar{A}) \rightarrow 0\}Q$	Causes z to be sent to $3S+0, 1, 2, 3$	3A; 0; 22, Z; Q
1 2A,	$\{0 \rightarrow D_0\}$	Places order n in D_0 in form $8n$	2A; 0; 12, Z
2 3A,	$\{(3) \rightarrow \bar{A}\}Q$	Clears (\bar{A})	3A; 3; 16, Z; Q
3 3A,	$\{(\bar{A}) \rightarrow 1\}Q$	Clears p_r store	3A; 1; 22, Z; Q
4 3A,	$\{(\bar{A}) \rightarrow 2\}Q$	Clears p_r store	3A; 2; 22, Z; Q
5 3A,	$\{(1) \rightarrow \bar{A}\}Q$	Sets p_{r-1} to \bar{A}	3A; 1; 16, Z; Q
6 3A,	$\{(0) \times \bar{A}\}Q$	Forms $p_{r-1}z$	3A; 0; 19, Z; Q
7	$\{(D_0) \pm \bar{K}\}$		0; 14, Z
8 3A,	$\{(4) \pm \bar{A}\}Q$	Adds a_r to (\bar{A})	3A; 4; 17, Z; Q
9	$\{1 \rightarrow D_0\}$		8; 13, Z
10 1A,	$\{12 \rightarrow \bar{S}\}$		1A; 12; 9, Z
11	$\{(\bar{L}) \rightarrow \bar{S}\}$	$(\bar{A})' = p = f(z)$	0; 11, Z
12 3A,	$\{(\bar{A}) \rightarrow 1\}Q$	Substitutes p_r in place of p_{r-1}	3A; 1; 22, Z; Q
13 3A,	$\{(2) \rightarrow \bar{A}\}Q$	Sets p'_r to \bar{A}	3A; 2; 16, Z; Q
14 3A,	$\{(0) \times \bar{A}\}Q$	Forms $p'_r z$	3A; 0; 19, Z; Q
15 3A,	$\{(1) \pm \bar{A}\}Q$	Adds p_r	3A; 1; 17, Z; Q
16 3A,	$\{(\bar{A}) \rightarrow 2\}Q$	Replaces p'_r by p'_{r+1}	3A; 2; 22, Z; Q
17 1A,	$\{5 \rightarrow \bar{S}\}$	Returns hyper-control to 5	1A; 5; 9, Z

Notes

1. The four-word hyper-locations of the data block; 3A, 0-3, 4-7, 8-11, 12-15, hold z , p_{r-1} , p'_r , and zero respectively.
2. The coefficients a_r occupy hyper-locations 3A 16-19, 20-23, etc.
3. During insertion of this hyper-routine, 32 holds the head location of this hyper-routine (1S) and location 33 holds $8np_{11}$, where n is the degree of the polynomial, while location 34 holds the head location of the 3S-data block.

The hyper-routine designed to compute $f(z)$ and $f'(z)$ is shown in Table 3. It is assumed that upon entry of hyper-control into this hyper-routine $(\bar{A})=z$. Storage space is provided in a data block for the variable z , for the partial polynomial p_{r-1} and the derivative p'_r , and the number zero used for clearing \bar{A} upon entry into the hyper-routine. Hyper-command number 10 could be replaced by $\{p_1 \rightarrow S\}$, since single counts into the sequence register made interpretively advance hyper-control one extra step forward.

When used in hyper-programmes, the coefficients would be inserted from tape and the make-up of the tape would be :

- (i) Control H (including P and Q facilities for hyper-programme).
- (ii) nT (which changes insertion-command to leave space for control parameters).
- (iii) Interpretive programme (function block, directory, and any additional function routines) starting at location n .
- (iv) $3S, x'T$ (which leaves space for data to be inserted subsequently and stores the head location of this block).
- (v) Input hyper-routine (outlined in Section XII) starting at x' .
- (vi) $4S$, input programme ; uses input hyper-routine and returns to primary B after $n+4$ items of input.
- (vii) $4A, 0 \rightarrow S, D$ (transfers to input programme to insert data following).
- (viii) Four zeros, clearing working space, and coefficients a_r in decimal code.
- (ix) $x'T$ (prepares to over-write hyper-programme just used).
($c(A) \rightarrow 33, D, 8np_{11}, U$) sets parameter for degree of polynomial.
- (x) Polynomial hyper-routine (reference to $3S$ -ed data block now possible).
- (xi) xS master hyper-programme (varies with problem).
- (xii) $xA, 0 \rightarrow S, D$ (transfer control to hyper-programme and starts computation).

XVII.—OTHER HYPER-ADDRESS CODES

The great flexibility of the interpretive method is obtained at the expense of time, one hyper-command corresponding in time to a full cycle of the interpretive loop and any function selected from the function block. Usually, in the systems adopted, the time spent in the function block performing essentially difficult and involved operations cannot be avoided and is considerably greater than the time occupied in one traverse of the interpretation routine.

However, as the length of a hyper-word becomes greater, more time is spent in transfers to and from the hyper-accumulator. A change of the hyper-code to a three-address code avoids much of this extra organizational shifting of data. This method is adopted in the n -fold accuracy function block. In this system there is no hyper-accumulator. One hyper-command occupies two store locations. The first contains a numerical address $m_n^{(1)}$, an operation code F , say, and the distinguishing Z . The following location contains the second and third numerical addresses, $m_n^{(2)}$ and $m_n^{(3)}$, say, in p_1 - p_{10} and p_{11} - p_{20} groups. Single address words without Z are interpreted as normal machine operations. Thus, for instance, the two-word hyper-command coded as $m_n^{(1)}$; $F, Z: m_n^{(2)}$; $m_n^{(3)}$ may cause $(m_n^{(1)})$ to be added to $(m_n^{(2)})$, each n words long, and placed into the n locations $m_n^{(3)}$, and so on according to the code number of F .

When, as occurs with large matrices, there are relatively few groups of data in the store, the address digits of one-word hyper-commands may be partitioned into $a, b, c; F, Z$. Here a, b, c , have values 0-8 (3 digits each) and refer to two of 16 store locations. Each of these groups of two words may be considered as partitioned into N, n, m, l , where N is the location of the first word in an $n \times m$ matrix of p -word numbers occupying l store locations in all.

The associated function block can be designed to provide operations such as matrix multiplication, inversion, division, etc. Additional calculations are performed, for each operation, to adjust the value of N , n , m , l of the result. Thus, in the multiplication which we may denote by $\{(a).(b) \rightarrow c\}$, $n_c = m_a$, $m_c = n_b$, and $l_c = p \times n_c \times m_c$ are calculated in the course of the operation and N_c may be provided by the programmer. By suitable design of the function block, the computer can be made to keep a record of usage of store space and adjust N_c itself. In these ways the programmer may be relieved of concern for lengths of data units and may programme in terms of a, b, c only.

The interpretive system has so far been applied mostly to various arithmetical methods as described, but has great potentialities for problems other than those that are purely arithmetical, since hyper-commands may be partitioned and coupled together in any manner desired so long as a suitable interpretation routine can be designed.

XVIII. REFERENCES

- PEARCEY, T., and HILL, G. W. (1953a).—Programme design for the C.S.I.R.O. Mark I Computer. I. *Aust. J. Phys.* **6**: 316-34.
- PEARCEY, T., and HILL, G. W. (1953b).—Programme design for the C.S.I.R.O. Mark I Computer. II. *Aust. J. Phys.* **6**: 335-56.
- WILKES, M. V., WHEELER, D. J., and GILL, S. (1951).—"The Preparation of Programmes for an Electronic Digital Computer." pp. 162-4. (Addison-Wesley Press Inc.: Cambridge, Mass.)

THE EFFECT OF INTERPRETIVE TECHNIQUES ON FUNCTIONAL DESIGN OF COMPUTERS

By T. PEARCEY,* G. W. HILL,* and R. D. RYAN*

[*Manuscript received November 12, 1953*]

Summary

An analysis of the programmes for a number of computations performed by the C.S.I.R.O. Mark I computer shows that a great proportion of store space and operating time is occupied by the control of the course of the calculation, and that the proportion of the store space required as working space by a programme decreases as the size of the programme increases. The increasing use and great flexibility of interpretive techniques suggests that a very flexible, reliable computer, easy to use, could be constructed. Such a computer would possess only a relatively small amount of rapid-access erasable store, and a larger amount of rapid-access non-erasable store, in which would be held all interpretation routines, function blocks, and so on. The operator would require no knowledge of the actual machine code, but would place his hyper-programmes and data into a slow-speed backing store.

I. INTRODUCTION

In a previous paper (Pearcey and Hill 1954) a very powerful technique was described for the simplification of programme design in cases of complicated arithmetical computations. This is known as the interpretive technique and was described in detail as applied to the C.S.I.R.O. Mark I computer. The method is also applicable to a wide range of programmes and not necessarily restricted to essentially arithmetical computations.

The effect of the interpretive technique, so far as the user is concerned, is to provide the computer with a number of additional and convenient functions which are, to his view, built into the machine. Further, this set of functions can be changed by use of different function blocks.

The number of machine commands in the function blocks is frequently large, from 300 to 700 words or more depending upon the particular hyper-functions desired. Hyper-programmes are more frequently of the order of 100 hyper-words only.

An analysis of the interpretive function blocks used in the C.S.I.R.O. Mark I computer shows that less than 5 per cent. of the space occupied by an interpretive programme and its auxiliary routines is subject to change as the calculation proceeds. This estimate includes working space, variable commands and control parameters, link storage, and so on.

This indicates that a new form of storage of very simple design, little of which need be erasable, could be used. Such a computer would operate entirely

* Division of Radiophysics, C.S.I.R.O., University Grounds, Sydney.

interpretively and possess a simple machine code, amounting at most to only 15 single operations. The user would require no knowledge of the machine code, but only of the hyper-code suitable for his computation. The programmer may choose a function block with a standard hyper-code or design one to suit the problem. This would greatly assist users to programme for "difficult computations".

Such a computer, effectively capable of doing very elaborate operations, could be constructed with less equipment than exists in most currently operating machines. This would considerably improve reliability, assist maintenance, and achieve a more even balance of the responsibility for designing and constructing such a computer between the engineer and the mathematician. Simplification and reduction of equipment relieves the engineer of much effort and places more responsibility on the shoulders of the mathematician, who designs the routines for what the user could consider as "built-in" functions.

II. ANALYSIS OF PROGRAMMES

The commands of a programme can be grouped according to type (transfer, sequence shifts, addition, etc.), or according to their purpose (computation, sequence control, etc.). The manner in which the store is used is indicated by the proportion of store space occupied by such groups. The relative frequency of performance of commands of such groups is a useful basis for estimating efficiency of use of time.

A number of widely different programmes and interpretive function blocks have been analysed in these ways. The problem programmes include programmes for X-ray crystallographic synthesis, experimental data analysis, integration of singular functions, evaluation of determinants, matrix operations, and solution of various types of partial differential equations. The interpretive systems included those for floating decimal arithmetic, double precision arithmetic, and an arithmetic for which the index point occurs between the p_{11} and p_{10} digits, for both the real and complex variable.

III. ANALYSIS OF STORE USAGE

This analysis corresponds to a simple count of the commands of different groups as they occur in written programmes. The results of grouping commands according to type are shown in Table 1 for the case of problem programmes and interpretive function blocks. All values are percentages of the total store space occupied by the programme, excluding command space required only during insertion of the programme into the computer. The standard deviation listed with each percentage is believed to be reliable to 1 per cent.

It will be noticed that a rather large proportion, 32 per cent., of the total consists of plain transfers. The small standard deviation of this group, 2-4 per cent., indicates that such transfers form an essential part of the programmes and could not be reduced below about 30 per cent., even with the address system of Mark I. The overall length of programmes in a one-address system, like that of EDSAC, shows an average increase of 25 per cent. over the store space occupied by equivalent programmes in the Mark I system.

Most of the increase is due to additional transfers required, which bring the proportion of store space occupied by transfers in a one-address system to about 40 per cent.

The frequency of "add" or "subtract" commands is seen to be greater in problem programmes than in interpretive function blocks. Greater inter-connection in function blocks increases the proportion of commands devoted to sequence shifts, which organize repetitions of the relatively smaller proportion

TABLE 1
STORE USAGE ANALYSIS: PERCENTAGE OF TOTAL PROGRAMME SPACE OCCUPIED BY COMMANDS OF VARIOUS TYPES

Programme	Type					
	Transfer	Add-Subtract	Control Shift	Discrimination	Variable	Other
Problem programmes	32±2	24±4	17±2	8±1	4±2	17±5
Interpretive function blocks	32±4	14±4	29±6	5±1	5±1	15±5

Length of problem programmes, 100-300 commands

Length of interpretive function blocks, 250-500 commands

of commands devoted to arithmetic operations. This emphasis, indicated by the proportion of sequence shifts, is found to increase with the degree of complexity of the arithmetical system.

In both types of programme, discriminations are relatively few in number, 5-8 per cent. Use of "E" and "G" type commands in the one-address EDSAC system for sequence shifts as well as for discriminations would bring the proportion of such operations to 20-30 per cent. It is very important to note that the

TABLE 2
STORE USAGE ANALYSIS: PERCENTAGE OF TOTAL COMMAND SPACE OF PROBLEM PROGRAMMES DEVOTED TO VARIOUS PURPOSES

Computation	43±16
Sequence control	23± 7
Linkage	15± 5
Command variation	13±10
Other	7± 1

proportion of variable commands is quite small, 4-5 per cent. Although this type of command occupies relatively little space, it is vitally necessary in flexible programme design.

The results of grouping commands in problem programmes according to purpose is shown in Table 2. This shows that less than half the programme space is devoted to calculation, and this includes all transfers etc. involved in

the actual calculation of results. Without transfers and similar organizational commands, the proportion devoted to calculation drops from 43 per cent. to about 30 per cent. Apart from the 7 per cent. devoted to connection with the user (input, output, hoots,* etc.) the remainder is devoted to controlling the course of the calculation.

IV. MACHINE SPEED EFFICIENCY

Since commands in some loops are repeated more often than others, a simple count of commands in written programmes is not a reliable indication of relative frequencies of performance of command groups. An analysis of performance frequencies shows considerably greater variation than the store space analysis. Depending on the programme, the proportion of time devoted to calculation of results varies from 30 to 70 per cent. The proportion devoted to organization varies from 70 to 30 per cent. Communication with the user by hoots, input, prints, etc. may occupy up to 50 per cent. of the operating time and is very variable. Of significance is the fact that multiplication occupies up to 25 per cent. of the time. In such cases, replacement of the machine operation by an equivalent routine for multiplication would decrease multiplication speed by a factor of about 240 and overall computing speed would be reduced by, at most, a factor of about 60.

V. VARIABLES AND WORKING SPACE

Only a small proportion of a programme changes during operation, and the working space it requires is only a small proportion of its size in the store.

Programmes require working space for the data operated on (e.g. accumulators, temporary storage space, etc.), store space to hold constants, parameters, variable commands, count parameters, and other cycling controls. All locations occupied by data of these types must be erasable, since the contents change both during calculation and from programme to programme. In general the proportion of programme space occupied by such variables tends to decrease as the size of the programme increases, due to use of the same stores for variables of many routines.

While single routines from the Mark I library, with lengths 30–90 commands, required 13–17 per cent. of their lengths for constants and working and variable spaces, problem programmes consisting of groups of library and special routines required only 4–10 per cent. of their overall length of 180–350 commands. This drop is largely due to use of the same working space for all routines, e.g. the working registers *A*, *B*, *C*, *D*, and *H*.

Problem programmes rarely exceed 350–400 commands in length. Problems requiring large groups of commands have been those suited to interpretive techniques, involving function blocks of lengths 300 commands upwards. The hyper-routines and hyper-programmes range in length from 100 commands upwards, but the amount of variable store required by the function block is independent of this further length. In estimating the variable space required, all hyper-registers (\bar{A} , \bar{X} , \bar{S} , \bar{K} , etc.), temporary storage spaces, and link and

* A "hoot" is a signal made audible by the loudspeaker.

loop control spaces must be included. About 20 locations are required for "control" registers, and a further $5n$ locations for "arithmetical" registers for n -word numbers. In cases of complex variables, this latter contribution is double that for real variables. Any increase in the effective value of n increases the total size of the function block (" $n=4$ " block is 40-60 per cent. longer than " $n=2$ " block), thereby tending to keep constant the ratio of variable space to programme space. For interpretive function blocks of lengths ranging from 250 to 600 commands used with the Mark I computer, this ratio was 4.5-5.5 per cent.

As the value of n increases, a stage is reached where it is more efficient to change from a one-address system to a three-address system. This avoids waste of store space for n -word arithmetical registers and waste of time for n -word transfers. Eliminating hyper-arithmetical registers causes the proportion of variable stores to decrease below 5 per cent. with increase in size of the function block. This suggests that there is an upper limit of about 100 store spaces to the amount of variable store needed by all types of programmes.

Experience so far shows a greatly increasing use of interpretive systems in Mark I, with frequent use of the same function blocks. The general utility, simplicity, and flexibility of interpretive methods tends to make them popular with users.

Computers, suitably designed for use of interpretive methods, could store the standard function block in a fixed or semi-fixed storage system of rapid access; only a relatively small additional amount of rapid-access erasable store being required for working space, etc. Large amounts of problem data would normally be held in a slower-access backing store since such stores of large capacity are inexpensive. Small amounts of problem data may be held also in more expensive rapid-access stores.

VI. OPERATION RATES IN HYPER-CONTROL

The advantages of interpretive methods are obtained at the expense of reduced speed of hyper-operations. Hyper-speed, the rate at which hyper-commands are performed, is considerably slower than machine-command speed and varies with the design of the function block used. A measure of the hyper-speed is the number of machine commands performed by each hyper-operation. This is listed in Table 3 for each hyper-function of a number of interpretive systems used in the Mark I computer. These values may be reduced to time in seconds for the Mark I computer by division by 500. The command symbolism for hyper-functions is that given in a previous paper (Pearcey and Hill 1954).

It will be noted that machine commands, those hyper-commands not terminated by Z , are performed most rapidly (14-18 command times). Transfers to and from the accumulator, real addition and subtraction, hyper-sequence and hyper-count operations are relatively rapid (40-60 commands). Multiplication hyper rate varies from 50 to 100 commands for the real variable. Hyper rate for division and square rooting decreases from 100-300 commands for single word precision to about 1500 commands for extended precision, due to organizational complexities introduced by extension to high precision. To the input

and output rates of 100–1000 commands must be added the time required for the mechanical operations of reading, printing, or punching.

For complex variables the hyper rate of transfers is only slightly decreased, that of addition and subtraction is roughly halved and of multiplication, division, and square rooting is decreased by four to five times.

The average rate of performance of any hyper-programme depends on the particular hyper-commands of which it consists. An analysis of the frequency of use of hyper-commands has been made for hyper-programmes for solution

TABLE 3
HYPER-SPEEDS

Real Hyper-operations*	No. of Machine Commands† per Hyper-operation				Complex Hyper-operations	No. of Machine Commands per Hyper-operation			
	A	B	C	D		A	B	C	D
$(n_p) \rightarrow \bar{A}$	40	40	43	50	$(n_q) \rightarrow \bar{A}$	50	50	44	59
$(n_p) \rightarrow \bar{A}$	43	72	50	160	$(n_q) \rightarrow \bar{A}$	90	180	57	400
$(n_p) \rightarrow \bar{A}$	55	72	51	180	$(n_q) \rightarrow \bar{A}$	140	180	61	420
$(n_p) \rightarrow \bar{A}$	104	57	51	150	$(n_q) \rightarrow \bar{A}$	430	350	77	1100
$(n_p) \rightarrow \bar{A}$	1500‡	120‡	123‡	1400‡	$(n_q) \rightarrow \bar{A}$	4700‡	750‡	280‡	4600‡
$(n_p) \rightarrow \bar{A}$	1500‡	260‡	250‡	1700‡	$(n_q) \rightarrow \bar{A}$	5000‡	1400‡	670‡	6500‡
$(n_p) \rightarrow 0t$	930§	180§	100§	980§	$(n_q) \rightarrow 0t$	1850§	300§	170§	1900§
$(I) \rightarrow n_p$	240§	190§	130§	320§	$(I) \rightarrow n_q$	470§	320§	220§	600§
$(\bar{A}) \rightarrow n_p$	40	40	46	50	$(\bar{A}) \rightarrow n_q$	52	52	52	66
$n \rightarrow \bar{S}$	34	35	40	42	$ n_q ^2 \rightarrow \bar{A}_p$	220	170	61	540
$n \rightarrow \bar{L}$	40	41	46	48	$(n_p) \rightarrow \bar{A}_q$	250	130	79	390
$(\bar{L}) \rightarrow \bar{S}$	37	38	43	45	$(n_p) \rightarrow \bar{A}_p$	3500‡	280‡	220‡	3000‡
$n_p \rightarrow D_r$	45	46	51	53	$n_q \rightarrow D_r$	45	46	51	53
$n_p \rightarrow D_r$	44	45	50	52	$n_q \rightarrow D_r$	44	45	50	52
$(D_r) \rightarrow \bar{K}$	43	44	49	51					
$f(A) \rightarrow \bar{A}$	60‡	60‡	65‡	65‡					
Machine command	14	14	16	18					

* The command symbols are those described by Pearcey and Hill (1954).

† The arithmetical systems involved in the columns of hyper-rates are : A, double precision arithmetic ; B, floating decimal arithmetic ; C, arithmetic with index point between p_{11} , p_{10} digits ; D, floating decimal index, double precision arithmetic.

‡ Average values only.

§ Not including mechanical printing or reading time.

of polynomial equations using complex arithmetic, integration of singular functions using floating index arithmetic, and matrix operations in floating index, double precision arithmetic.

Of all the hyper-commands performed, hyper-transfers form 10–40 per cent., hyper-control operations 10–40 per cent., addition and subtraction 5–20 per cent., multiplication 5–20 per cent., and machine commands less than 30 per cent., while the relatively slower operations, division and square rooting, require less than 5 per cent., with input and output less than 5 per cent.

By suitably combining the hyper rates with the frequencies of performance of hyper-commands, an average hyper rate of 50-90 is obtained for the cases cited. For an estimate of efficiency these figures must be compared with the average speed with which such operations could be performed by the direct technique of cueing and linking outlined in a previous paper (Pearcey and Hill, 1953). By direct techniques the duration of each hyper-operation could be decreased by omission of the 30 "redundant" machine commands in the interpretation loop reducing the equivalent operation speed to 20-60 machine commands per hyper-operation. At the same time most hyper-commands would be replaced by 2-3 machine commands for planting address, link datum, and operation code, thus doubling or trebling the length of the hyper-programme. Thus interpretive techniques are seen to be roughly twice as expensive in time and half as expensive in command space as direct programme techniques.

VII. REFERENCE TO SLOW-ACCESS STORE

If problem data and hyper-programme are stored in a slow-access store such as a magnetic drum, operating speed would be seriously reduced if transfers to and from the slow-access store were very frequent. While all the hyper-operations would require a transfer of the hyper-command from the slow-access store, a proportion p involve a further transfer of an n -word number. In an interpretive system for n -word numbers, with an average hyper rate of r machine commands performed for every hyper-command, some $r/(1+pn)$ machine commands would be performed on the average for very slow-access store transfer.

In the cases cited, $r/(1+pn)=15, 20, 50$, so that, unless the access time of the low-speed store were more than 50 times that of the rapid-access command store, there would be no serious reduction in operating speed. The store access ratio (ratio of low-speed to high-speed access times) should not be greater than 50 and preferably less than 15.

VIII. BEARING UPON FUNCTIONAL DESIGN

The results of these discussions have direct bearing upon the functional design of computers. The factors which are seen to be important are as follows:

- (1) The interpretive method simplifies programme design and has the effect of providing the user with many additional "built-in" functions.
- (2) The interpretive method is flexible, the code and address style being chosen at will.
- (3) Function blocks can be designed for interpretive programmes which will be of very common use. A small number of standard blocks would satisfy the great majority of users.
- (4) The proportion of the store required by a programme as working space decreases as the size of the programme increases. For interpretive arithmetical programmes this is about 5 per cent. of the programme space only. This only need be erasable during operation, apart from the problem data store.
- (5) Considerable time is spent by transfers and control operations and in the control of the programme itself.

(6) Reference to data for computation is relatively infrequent. Such data could be held in a slow-access store with only a small degree of loss of time.

(7) About 50-90 machine commands are performed to each hyper-command.

Accepting these points, a computer could be designed having adequate speed, relatively simple engineering requirements, and great flexibility and ease in programming. Thus we may adopt the following principles for its design :

(1) Use the interpretive method of programming entirely.

(2) The machine code should be short and possess only a small number of simple functions, e.g. a one-address system.

(3) The user would use only a hyper-code and would not require knowledge of the machine code.

(4) Commands should be adopted from a rapid-access store, data and hyper-programmes being held in a slower access or backing store.

(5) The rapid-access store need have only a small fraction of its total capacity erasable, the rest fixed, since function blocks would be stored permanently. The erasable store would be used as working space for the function block and its associated interpretation routine and directory. Provision should be made for adopting variable commands from the erasable part of the store.

From this it will be seen that the hyper-programme and problem data, the parts supplied by the user, would normally be held in a backing store only. The machine adopts its commands from the rapid-access store. The fact that only a small part of the high-speed store need be erasable should considerably simplify the engineering problems associated with rapid-access storage.

IX. THE STORE

As in the case of most computers, the logical design depends largely upon the physical nature of the storage system adopted. Much effort is expended in the design of large-capacity storage systems possessing write-erase features and rapid accessibility. This expenditure could be avoided with probable reduction of equipment and additional reliability if a small amount of rapid-access erasable store were used together with a larger amount of fixed or, rather, semi-permanent store. The erasable store would provide working space and all hyper-registers, and the function blocks providing the hyper-functions would be inserted into the non-erasable store. These stores will be called the variable and fixed high-speed stores. The serial mode of operation will be presumed on the grounds that less equipment is required in the control and operation of such a computer than in an equivalent parallel operating machine. Digit recurrence rates of at least one megacycle should be attainable.

It is also clear that a large backing store must be provided for storing large amounts of erasable problem data and the hyper-programme and hyper-routines.

X. FIXED HIGH-SPEED STORE

For the fixed store a "flying spot" system similar to that used in the transmission of films by television seems a possible technique, not requiring excessive effort for development. In this system the image on the screen of

a high definition cathode-ray tube is focused upon a film frame which is scanned by a light spot tracing a raster on the face of the screen. The light transmitted by the film is detected for transmission by a photomultiplier tube.

By replacing the film by a matrix representing the routines of the function block and other standard routines, and controlling the passage and position of the light spot on the screen, the output of the photomultiplier tube may be made to provide commands to the computer.

The advantages of such a system would be those of relatively small amounts of standard equipment, high digit capacity per frame of the matrix, and small access time equal to the time required to shift the spot from one place to another.

Commands and standard constants could be stored on the matrix in rows of spots in binary code. It seems possible to store up to one thousand 16-digit commands on one matrix. The application of suitable deflexion voltages corresponding to a given serial number would position the spot ready to read out a selected command. The spot may then be brightened and moved across the screen at a fixed rate so as to scan the required word on the matrix. Such a scheme would provide serial transmission of digits, with a possible digit rate of one megacycle.

In practice it may be possible to avoid the use of an optical system by placing the matrix directly against the cathode-ray oscilloscope face. The store could be extended by addition of similar units.

The disadvantage of the fixed store is that special matrices must be constructed either photographically or mechanically. Against this argument, however, such matrices would be standard, few in number (about 20 different matrices), made once only, and frequently used.

Other methods of constructing stores involve the use of one element for each digit as in the case of magnetic cores or dielectric elements. With the former of these a considerable reduction of the number of cores can be attained by suitably weaving a number of reading circuits through one row of cores. However, for convenience of rapid changing the optical system is preferable.

The use of fixed store matrices would relieve the programmer of much tiresome work in the use of multiple accuracy and other elaborate arithmetical methods.

XI. ERASABLE HIGH-SPEED STORE

This store could take one of a number of possible forms, such as magnetic core or electrostatic matrices, delay lines, etc. Delay lines, although "volatile", may be made to operate at higher than one megacycle recurrence rates and would be highly suitable for serial operation. The access time would be reduced by having as few words as possible placed in each delay line, consistent with there being a reasonably small amount of equipment. One possible means of achieving this is to adopt a "multiplex" scheme in which more than one word is held in a delay line of one word "length" by interspacing the digits of the words stored. The access time to any word is thus reduced to a maximum of one word-time.

The digit period of both types of store must be the same since the transfers are presumed to be serial. The storage locations may be numbered serially (0 to $N-1$), the erasable store occupying the earliest positions (0 to $M-1$), with the fixed store following without a break (M to $N-1$).

XII. LOW-SPEED STORE

This will be referred to by the programme held in the high-speed store and will thus not be required to possess an access time as small as that needed in the high-speed store. Nevertheless, the access time for the backing store, or low-speed store, should be as small as possible within reasonable engineering safety margins. This store would probably take the form of a magnetic drum, a system which is known to be highly reliable and economical in space and equipment.

XIII. THE CODE SYSTEM

Commands of the fixed store matrices are recorded in "machine code", each code number corresponding to a single machine function. This code is fixed by the design of the machine. The interpretation codes are variable and are chosen by the user or the mathematician to suit the type of calculation. To each of the different interpretation codes there will correspond a set of machine code matrices. It will be assumed that the machine operates in the binary code, with complementary representation of negative numbers, the sign digit being the last digit transferred in each word.

XIV. THE MACHINE CODE

From the point of view of simplicity of use, of logical design, and of minimum equipment the best design would involve serial operation in a one-address code system with only essential machine functions. The consequent disadvantages of increase in programme size and reduction of machine speed is amply compensated by the size of the cheap store available and its reduced access time.

Each high-speed store location will be referred to by its address n , and its content denoted by (n) . The total number of low-speed store locations will be N' and any particular such location will be referred to by its address n' with content (n') .

We shall assume that one word occupies 16 binary digits, although this may be chosen at will, and that all transfers called by the machine code will be 16-digit transfers. There would be no basic change to make such a machine to possess a 16-digit command code and to transfer 32 digits under such a code.

A central accumulator A of double length capacity, i.e. 32 digits for 16-digit transfers, whose content is denoted by (A) , must be provided. The accumulator could be of the delay line type with suitable input and output gates and special devices for multiplication and left shifts. Though multiplication could be performed by a programme stored in the machine by repeated addition and shifting, it is so frequently used in practice that it is best to provide built-in multiplication. In this case additional arithmetical registers must be provided but need not be referred to explicitly by the machine code.

Operation proceeds serially, that is, commands are adopted from sequential high-speed store locations except when transfers of control are called. The

sequence register containing the current command address will be referred to as S and its content as (S) . The sequence register must be provided with a $\frac{1}{2}$ -adder so that it may generate sequential addresses by addition of a unit digit from the sequence unit after extraction of each command. These facilities would also be used whenever sign test functions are called. The interpreter, which receives each command for decoding, will be denoted by K . The input register I and output register O connect the machine with the operator.

All commands will contain a store address and a function number. Some of these functions will refer to variable store $0 \leq n < M$, to the fixed store $M \leq n < N$, and to the backing store $0 \leq n' < N'$, and we want to allow N' to be as great as possible.

It is found that there are only three commands which refer to fixed store locations and to low-speed store locations. These are :

- | | | |
|--|------------|----------------------|
| (1) Shift control to n | denoted by | $n \rightarrow S$ |
| (2) Transfer (A) to n' in
the low-speed store | „ „ | $(A) \rightarrow n'$ |
| (3) Transfer (n') in the
low-speed store to A | „ „ | $(n') \rightarrow A$ |

A special group of two digits in the function part of the code may be used to distinguish these from all other functions. These may be the two most significant digits, i.e. p_{16} and p_{15} of a 16-digit command consisting of digits $p_{16}, p_{15}, \dots, p_1$ in descending order of significance.

Other commands refer only to the erasable high-speed store $0 \leq n < M$, for which fewer address digits are required, e.g. p_1, \dots, p_8 provides for $0 \leq n < 256$.

Hence the following address scheme may be adopted :

p_{16}	p_{15}	$p_{14}-p_9$	p_8-p_1	Function
1	$1 \leftarrow 0 \leq n' < N'$	—————→		$(n') \rightarrow A$
1	$0 \leftarrow 0 \leq n' < N'$	—————→		$(A) \rightarrow n'$
0	$1 \leftarrow 0 \leq n < N$	—————→		$n \rightarrow S$
0	$0 \leftarrow \text{Function code} \rightarrow \leftarrow 0 \leq n < M \rightarrow$			Other functions

As illustrated, N and N' would be limited to 16,384 locations. This is probably excessive for N .

The remaining functions may be chosen as follows and refer only to $0 \leq n < M$:

- | | | |
|---|------------|----------------------|
| (1) Transfer (n) to A (accumulator) | denoted by | $(n) \rightarrow A$ |
| (2) Transfer (A) to n | „ „ | $(A) \rightarrow n$ |
| (3) Add (n) to (A) and hold the sum in A | „ „ | $(n) \dot{+} A$ |
| (4) Subtract (n) from (A) and hold the
difference in A | „ „ | $(n) \dot{-} A$ |
| (5) Form digit by digit product of (n) with
regard to (A) and place in A | „ „ | $(n) \dot{\times} A$ |

(6) Form product of (n) and (A) and place in A	„ „	$(n) \times A$
(7) Rotate (A) by n places to the left	„ „	$n \rightarrow L$
(8) If (A) is -ve add unit to S (sequence register)	„ „	$S(A) \xrightarrow{c} S$
(9) Transfer content of input register (I) to location n	„ „	$(I) \rightarrow n$
(10) Transfer (n) to output register O and record	„ „	$(n) \rightarrow O$
(11) Transfer (S) to location n	„ „	$(S) \rightarrow n$
(12) Transfer (n) to S	„ „	$(n) \rightarrow S$
(13) If (n) non-zero stop sequence	„ „	$(n) \rightarrow T$
(14) Transfer (n) to the interpreter K and add the next command to it	„ „	$(n) \rightarrow K$

Of these commands, (5) is useful in any interpretation routine and will therefore be used in all programmes, and (7) allows of both the greater and lesser half of products to be transferred from A and also facilitates the use of strobe methods in routines.

Command (8) allows all discriminations to be made, with the aid of the shift command (7) on any digit not necessarily the sign digit. Commands (11) and (12) allow links to be stored in the erasable store and hence allow transfers to and from routines in the fixed store to be made. Command (14) is special and allows for variable commands. The address may be stored as a parameter in the erasable store at n . If $(n) \rightarrow K$ is placed in m and $(A) \rightarrow 4$ is placed in $m+1$, then the latter command is adopted as $(A) \rightarrow 4+(n)$. It will be noticed that machine commands may be adopted from either the fixed or erasable stores and that the serial ordering of the locations of the erasable part continues into the fixed store, some locations of which may exist for which $n < M$. Standard data used by the function blocks will lie either in the erasable store or in that part of the fixed store for which $n < M$.

XV. STORE SELECTORS

Considerable convenience is achieved by adopting commands from the erasable store as well as from the fixed store. This would assist in insertion of hyper-programmes and data by enabling the insertion routine to be held in the erasable store. Further, short programmes of standard type or short hyper-programmes could be stored in the erasable part.

There would be three store selector units, each of a different type; one for the erasable store, one for the cathode-ray oscilloscope fixed store, and one for the backing store. In selection of commands the contents of the sequence register would be transmitted to two selectors, that for the erasable store and that for the fixed store. If $0 \leq (S) < M$, the p_{14} - p_9 digits would be zero and p_8 - p_1 digits in the erasable store selector would operate a "tree-type" selector preparing the erasable store to transmit. For $M \leq (S) < N$, the p_{14} - p_{11} digits (p_{14} - p_9 digits not all zero) in the fixed-store selector would apply the required

deflection potential for placing the spot on the cathode-ray oscilloscope at the beginning of the required trace. The digit group $p_{14}-p_9$ thus selects which part of the high-speed store is used ; if this group is clear of units the erasable store will be operated ; otherwise the fixed store would be used.

The backing-store selector could be of the counting-coincidence-selector type operated by digits p_1-p_{14} ($0 \leq n < N'$) and would be used in selection of locations for transmission of data or hyper-commands during performance of machine commands.

XVI. THE HYPER-CODES

Hyper-functions will vary and be chosen at will, and for each set suitable matrices will be designed.

In particular, three hyper-functions may refer to addresses n' in the low-speed store where $0 \leq n' < N'$. These will be :

- | | |
|---|---------------------------------------|
| (1) Transfer hyper-control to n' | denoted by $(n') \rightarrow \bar{S}$ |
| (2) Transfer the p -fold hyper-word in locations n' to $\overline{n+p-1}'$ to the hyper-accumulator | „ „ $(n_p') \rightarrow \bar{A}$ |
| (3) Transfer the content of the p -fold hyper-accumulator to the p -fold hyper-word in location n' to $\overline{n+p-1}'$ | „ „ $(\bar{A}) \rightarrow n_p'$ |

In these functions p will be implicit and specified in the routines of the fixed store (see Pearcey and Hill 1954). All other hyper-functions will refer at most to the M locations of the erasable store. In such cases, $n (< M)$ must be specified in the hyper-command, for instance, in digit positions p_1 to p_8 , leaving digits p_9 to p_{14} for the remaining hyper-function code providing for 64 possible hyper-functions.

By suitably choosing the hyper-code it may be possible to extend the code to additional hyper-functions and to change from one to another mode of interpretation, packing more than one hyper-command into a single word, or extending to a two- or three-address hyper-code as desired.

XVII. ORGANIZATION OF THE COMPUTER

The computer would contain only the essential registers. The cycle of operations would be as follows :

(1) The content of the sequence register is transmitted to the high-speed store selectors.

(2) The selected command is transmitted from the high-speed store to the interpreter register where it is decoded, and a unit is added to the sequence register.

(3) The numerical address held by the interpreter is transmitted to the high-speed store selector and to the low-speed store selector and the function addresses decoded by suitable selectors and suitable gating operations.

(4) The required transfer between registers takes place.

In the first of these operations the sequence register content is transferred to both fixed store and erasable store selectors. If the p_9 - p_{14} digit group is zero, the command is selected from the erasable store; otherwise from the fixed store. In the second operation of the computer sequence, the interpreter receives the command. The transfer occurs over a minor cycle period coincident with digit periods p_1 - p_{16} .

Decoding would occur in three parts. The interpreter may be thought of as a distributed constant delay line and similar to the sequence register. Digits p_{15} - p_{16} are connected to a separate decoder providing four outputs. Digits p_9 - p_{14} are connected to a "function decoder" of the tree type. Digits p_1 - p_{14} are transmitted in parallel to the variable store selectors and to the backing store selectors by activation of gates by the sequence unit waveforms.

Transfers from the sequence and interpreter registers could be made in parallel by initially tapping along the lines for the output digits.

For two of the configurations of p_{15} , p_{16} digits (1,1 and 0,1) the backing store output or input gate is activated, the store location being selected by the p_1 - p_{14} digits in the backing store selector. If the p_{15} , p_{16} digits are 1,0 respectively, the p_1 - p_{14} digits of the interpreter are gated in parallel to the sequence register. the configuration 0,0 of p_{15} , p_{16} digits permits the p_9 - p_{14} decoder to be activated, selecting the required function by opening gates for transferring the p_1 - p_8 digits in the erasable store selector to select the datum location.

Thus, in the first operation of the computer-sequence, either erasable or fixed stores may be called to transmit, and in the third operation either the backing store or erasable store is involved. In the case of low-speed backing store transfers, the sequence unit must not stimulate action until coincidence is found by the selector.

XVIII. SPEED OF OPERATION

The speed of operation obtained depends upon the access time of both rapid and slow stores. It seems reasonable to choose these speeds such that the access time of the low-speed store is about 20-40 times the access time to commands, so that operating time is fairly evenly divided between high-speed commands and backing store transfers.

The one-address system suggested turns out to be similar to that adopted for EDSAC (Wilkes, Wheeler, and Gill 1951) and differs considerably from that of the C.S.I.R.O. Mark I computer. It may be expected, therefore, that the function blocks designed for the proposed machine would occupy 25 per cent. more words than those listed in Table 3. The estimate of high-speed commands to each backing store transfer, suitably weighted, would amount to about 30 high-speed commands.

If we put the access time of the backing store at about 4 msec (in practice the smaller this is the better) and assume that a further 4 msec be occupied by the 30 high-speed operations, we arrive at a period of about 133 μ sec per high-speed function. At a 1 megacycle recurrence rate, the two serial transfers would occupy 32 μ sec, leaving over 50 μ sec each for operations 2 and 4 in the

computer cycle which involves switching or transferring digits to selectors and raising the selectors to suitable marginal levels. These speeds appear readily attainable and might even be exceeded.

At such a speed of operation the proposed computer would be able to do floating index and double precision operations at about one-eighth of the speed now required by the Mark I to perform normal operations, and corresponds to an increase in real speed over the Mark I for similar operations by a factor of five.

XIX. REFERENCES

- PEARCEY, T., and HILL, G. W. (1953).—Programme design for the C.S.I.R.O. Mark I computer.
II. Programme techniques. *Aust. J. Phys.* **6**: 335-56.
- PEARCEY, T., and HILL, G. W. (1954).—Programme design for the C.S.I.R.O. Mark I computer.
III. Adaptation of routines for elaborate arithmetical operations. *Aust. J. Phys.* **7**: 485.
- WILKES, M. V., WHEELER, D. J., and GILL, S. (1951).—"The Preparation of Programmes for an Electronic Digital Computer." pp. 162-4. (Addison-Wesley Press Inc.: Cambridge, Mass.)

SHORT COMMUNICATIONS

THE CONTRIBUTION OF PHONONS TO THE THOMSON COEFFICIENT*

By P. G. KLEMENS†

An electric current of density \mathbf{j} flowing in a temperature gradient ∇T gives up heat in a reversible way; the amount of heat thus liberated per unit volume and unit time is given by

$$-\mu \mathbf{j} \nabla T, \quad \dots\dots\dots (1)$$

μ being the Thomson coefficient, which can be shown to be

$$\mu = \frac{T}{e} \frac{d}{dT} \left\{ \frac{1}{T} \left(\frac{K_1}{K_0} - \zeta \right) \right\}, \quad \dots\dots\dots (2)$$

where ζ is the Fermi energy, and the K 's are the usual transport coefficients as defined by Mott and Jones (1936, p. 306, equation (99)), or more generally by Wilson (1953, p. 305).

In addition to this purely electronic Thomson coefficient, there is a component due to the lattice waves, arising as follows.

If a current \mathbf{j} is set up, the phonons, interacting with the electrons, have their distribution altered from equilibrium with a tendency to acquire excess quasi-momentum in the direction of the current. There is a resulting heat flow \mathbf{Q} , proportional to \mathbf{j} , defining a contribution to the Peltier coefficient $\pi_p = Q/j$. Since the Thomson and Peltier coefficients are related by

$$\mu = -\frac{\pi}{T} + \frac{d\pi}{dT}, \quad \dots\dots\dots (3)$$

the additional contribution to the Thomson coefficient is

$$\mu_p = \frac{1}{j} \frac{dQ}{dT} - \frac{1}{T} \frac{Q}{j}. \quad \dots\dots\dots (4)$$

The electron distribution function can be written in the form

$$f = f^0 + \frac{\lambda \cdot \mathbf{k}}{KT} \frac{df^0}{d\epsilon}, \quad \dots\dots\dots (5)$$

λ being a vector characterizing the deviation from equilibrium. Similarly the phonon distribution can be expressed as

$$n = N + \frac{\lambda' \cdot \mathbf{q}}{KT} \frac{dN}{dx}, \quad \dots\dots\dots (6)$$

* Manuscript received February 25, 1954.

† Division of Physics, C.S.I.R.O., University Grounds, Sydney.

f° and N being the respective equilibrium distributions, \mathbf{k} and \mathbf{q} the wave-vectors, $\epsilon = (E - \zeta)/KT$ and $x = \hbar\omega/2\pi KT$, E and $\hbar\omega/2\pi$ being the energies of the electrons and phonons respectively. The electrons and phonons interact so as to conserve energy and total wave-vector. If there are no other interactions, the electrons and phonons are in quasi-equilibrium if $\lambda = \lambda'$, as pointed out elsewhere (Peierls 1930*a*, 1930*b*, 1932; Klemens 1951*b*). True equilibrium, stationary for any interaction, requires $\lambda = \lambda' = 0$, but in the absence of processes capable of obliterating excess momentum ($\Sigma \mathbf{k} + \Sigma \mathbf{q}$), the electron-phonon interactions will merely distribute this excess momentum in such a way that λ and λ' are equal and constant for all electron and phonon energies.

The current density corresponding to a given value of λ is easily found from (5) to be

$$\mathbf{j} = \frac{2\pi e N}{\hbar} \lambda, \quad \dots\dots\dots (7)$$

where N is the number of free electrons per unit volume. Similarly the heat current \mathbf{Q} carried by phonons is found from (6) to be

$$\mathbf{Q} = \lambda' S T \frac{2\pi}{3\hbar}, \quad \dots\dots\dots (8)$$

where S is the lattice specific heat per unit volume. If the electron-phonon interaction is much stronger than the phonon-phonon interaction, which is due to anharmonicities or the elastic scattering of phonons by lattice imperfections, λ and λ' become equal, and μ_p is obtained from (4), (7), and (8). It is thus found that

$$\mu_p = \frac{1}{3Ne} T \frac{dS}{dT}, \quad \dots\dots\dots (9)$$

More generally, if τ_p is the effective relaxation time for phonon-phonon interactions in the sense used in the theory of the lattice thermal conductivity (Klemens 1951*a*), and if τ_e is the relaxation time for phonons interacting with electrons, it is easily seen that

$$\lambda' = \lambda \frac{\tau_p}{\tau_e + \tau_p}, \quad \dots\dots\dots (10)$$

and (9) has to be multiplied by the additional factor $\tau_p/(\tau_e + \tau_p)$. If the τ 's do not have the same frequency dependence, the expression becomes

$$\mu_p = \frac{1}{3Ne} T \frac{d}{dT} \left\{ \int S(\omega) \frac{\tau_p}{\tau_e + \tau_p} d\omega \right\}, \quad \dots\dots\dots (11)$$

where $S(\omega)d\omega$ is the contribution towards the specific heat from frequencies ω , $d\omega$.

The specific heat of a solid of N' atoms per unit volume is $233N'K(T/\theta)^3$ at very low temperatures, so that from (9)

$$\mu_p = \frac{233N'}{9Ne} \left(\frac{T}{\theta} \right)^3 K. \quad \dots\dots\dots (12)$$

For a free-electron metal the electronic Thomson coefficient is $-\pi^2 K^2 T / e \zeta$. Thus $\mu_e \propto T$ and $\mu_p \propto T^3$. As an example of orders of magnitude, if $\zeta / K \sim 10^4$ °K, $\theta \sim 2 \times 10^2$ °K, and $N' \sim N$, then $|\mu_p| \sim |\mu_e|$ at about 20 °K, provided of course that $\tau_p > \tau_e$ at this temperature. Whether μ_p is of appreciable magnitude at any temperature depends upon the relative magnitudes of τ_p and τ_e at those temperatures where (12) is comparable with μ_e .

For a pure metal μ_e is not strictly proportional to T , since the constant of proportionality varies with the dominant scattering mechanism (static imperfections at lowest temperatures, lattice waves at higher temperatures). In such cases it may not be possible to identify the electronic component μ_e . However, in an alloy this difficulty is avoided. The alloy must be such that $\tau_e < \tau_p$; this is fulfilled below about 10 °K for well-annealed silver-palladium alloys (Kemp *et al.* 1954). Thus it seems that μ_p may be observable in this and similar cases.

In the case of semi-conductors $\tau_p \ll \tau_e$; hence

$$\mu_p \sim \frac{1}{3Ne} T \frac{d}{dT} \left(S \frac{\tau_p}{\tau_e} \right). \quad \dots\dots\dots (13)$$

While the factor τ_p / τ_e is small, so is N ; hence μ_p may be appreciable, and will in general be proportional to the lattice thermal conductivity.

References

- KEMP, W. R. G., KLEMENS, P. G., SREEDHAR, A. L., and WHITE, G. K. (1954).—*Proc. Phys. Soc. Lond.* A **67**: 728.
 KLEMENS, P. G. (1951a).—*Proc. Roy. Soc. A* **208**: 108.
 KLEMENS, P. G. (1951b).—*Proc. Phys. Soc. Lond.* A **64**: 1030.
 MOTT, N. T., and JONES, H. (1936).—"Properties of Metals and Alloys." (Oxford Univ. Press.)
 PEIERLS, R. (1930a).—*Ann. Phys. Lpz.* **4**: 121.
 PEIERLS, R. (1930b).—*Ann. Phys. Lpz.* **5**: 244.
 PEIERLS, R. (1932).—*Ann. Phys. Lpz.* **12**: 154.
 WILSON, A. H. (1953).—"Theory of Metals." (Cambridge Univ. Press.)

FLUCTUATIONS OF LONG-PERIOD ACCUMULATIONS OF DAILY RAINFALL AMOUNTS*

By J. NEUMANN†

Introduction

In a recent communication, Bowen (1953) notes the tendency for heavy falls of rain at stations widely scattered over the southern hemisphere to occur on certain dates rather than on others. Surprisingly enough, records for the British Isles show maxima of frequency of heavy rain on the dates, or close to the dates, of peaks for the southern hemisphere stations concerned. As for the

* Manuscript received April 14, 1954.

† Meteorological Service, Hakirya, Tel Aviv, Israel.

latter stations no evidence is found for progressive displacement in the timing of peaks with geographical location, and as the data for the British Isles indicate that some of the peaks occur on dates identical, or nearly identical, with those for the southern hemisphere stations, Bowen suggests that an extraterrestrial factor must be operative resulting in increased precipitation on certain dates. In particular, he suggests that fine meteoritic dust provides rain-forming nuclei when it falls into the cloud tops of the lower atmosphere. The meteoritic dust is believed to originate from meteor streams in the solar system through which the Earth passes about one month prior to the dates of the daily rainfall peaks. For further details of this hypothesis, reference is made to the original paper.

Problem of the Present Paper

On reading Bowen's paper, an examination of local rainfall data appeared to be worth while. In the archives of the Climatological Division of the Israel Meteorological Service, records of daily rainfall amounts have been readily available for a number of stations in the area of the former British Mandatory Palestine. For dates between January 1 and February 5,* peaks exceeding the mean by 50 per cent. appear on the dates cited in Table 1.

TABLE 1
DATES OF PEAKS OF ACCUMULATED DAILY RAINFALL FOR SOME STATIONS IN THE AREA OF THE
FORMER PALESTINE

Station	Geographical Coordinates		Years	Dates
	N.	E.		
Haifa	32° 49'	34° 59'	1880-1939	Jan. 10, 14, 21
Jenin	32° 28'	35° 18'	1921-1939	Jan. 8, 12-13, 30-31; Feb. 3-4
Tel Aviv City	32° 04'	34° 46'	1924-1950	Jan. 13, 20, 28; Feb. 2
Jerusalem, American Colony	31° 48'	35° 14'	1898-1939	Jan. 13-14, 19-20
Jericho	31° 51'	35° 27'	1922-1939	Jan. 11-13, 15, 31; Feb. 1-4
Hebron	31° 32'	35° 04'	1895-1914	Jan. 7, 10, 18-19, 22, 27
Beersheba	31° 14'	34° 47'	1921-1939	Jan. 15, 19-20, 29-30; Feb. 2, 4-5
Gaza	31° 31'	34° 27'	1919-1939	Jan. 9, 12, 14, 18, 29, 31

Some of the dates in Table 1 are surprisingly close to the dates in Bowen's (1953, p. 492) Table 1. Particularly good is the timing of the peaks for Tel Aviv and Sydney. The daily rainfall amounts for Tel Aviv between the dates January 1 and February 5, in the years 1924-1950, are shown in Figure 1 (*a*).

The question arises whether the coincidence, or near-coincidence, of peaks at stations widely scattered in the two hemispheres is due to a physical factor exerting its influence the world over, as suggested by Bowen, or is, perhaps, the outcome of random fluctuations. The latter possibility should not be ignored

* Most of Bowen's data refer to this period.

ab initio despite the simultaneity of some of the peaks at different places. In a somewhat restricted sense, the problem is whether for any of the stations, the peaks can or cannot be the result of *random accumulations* of rainfall on the dates concerned. It is this restricted problem which will be considered below.

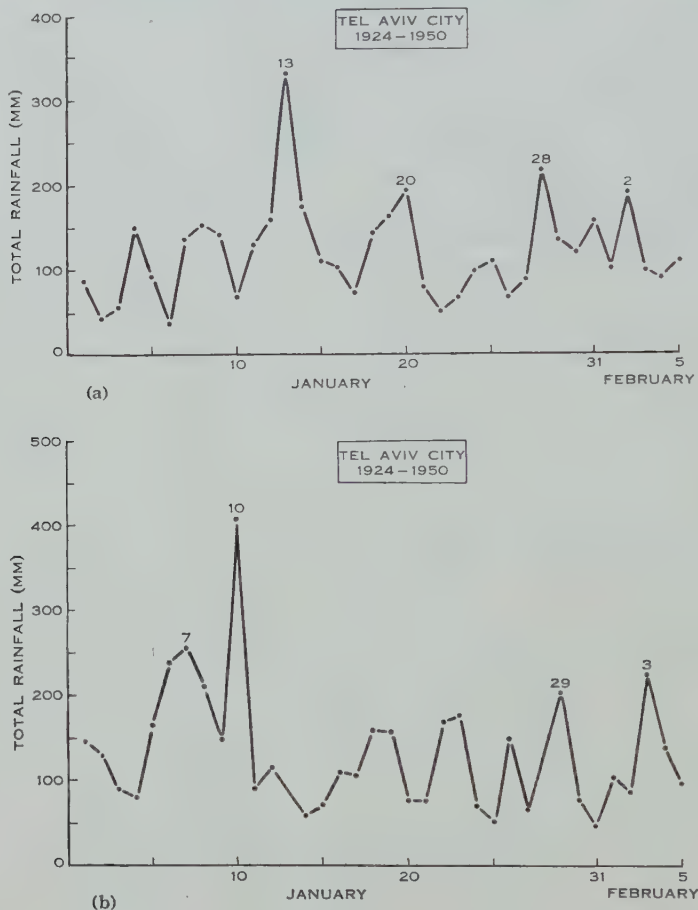


Fig. 1 (a).—Accumulated daily rainfall amounts for Tel Aviv City—original (true) data.

Fig. 1 (b).—"Accumulated daily rainfall amounts" for Tel Aviv City resulting from random shifting (in accordance with Table 2) of the original data.

Testing for Randomness

For examining the problem, a method devised by Bartels (1948) will in essence be followed. The idea underlying the method is as follows. Consider a time series the terms of which arise from accumulation of data and such that its peaks are believed to be produced by some physical factor becoming effective on certain dates. If now the basic data are reshuffled in some sort of random scheme, then in the series (accumulations) resulting from the reshuffling, the peaks should disappear. If, however, the new series be of a character similar

to the original and, e.g., peaks of comparable magnitude appear on other dates, then the original peaks are likely to be the result of random accumulations.

In the present investigation, the time series is the series of accumulated daily rainfall amounts. The physical factor the effect of which is supposed to become effective on certain dates is meteoritic dust. The method will be applied

TABLE 2
INSTRUCTIONS FOR THE RANDOM SHIFTING OF DAILY RAINFALL DATA
FOR THE TEL AVIV SERIES

Random numbers from a table published by Snedecor (1946, p. 12)

Year (from Snedecor's Rows)	Day (from Snedecor's Columns)	Column Marked "January 1"	
		For Year	Begins With
24	26	1924	Dec. 26, 1923
36	05	1936	Jan. 5, 1936
39	20	1939	Dec. 20, 1938
46	31	1946	Dec. 31, 1945
30	11	1930	Jan. 11, 1930
41	23	1941	Dec. 23, 1940
29	03	1929	Jan. 3, 1929
48	12	1948	Jan. 12, 1948
49	28	1949	Dec. 28, 1948
26	29	1926	Dec. 29, 1925
42	07	1942	Jan. 7, 1942
40	06	1940	Jan. 6, 1940
25	24	1925	Dec. 24, 1924
50	19	1950	Dec. 19, 1949
28	30	1928	Dec. 30, 1927
35	01	1935	Jan. 1, 1935
33	21	1933	Dec. 21, 1932
45	14	1945	Jan. 14, 1945
32	10	1932	Jan. 10, 1932
31	25	1931	Dec. 25, 1930
43	02	1943	Jan. 2, 1943
44	09	1944	Jan. 9, 1944
37	27	1937	Dec. 27, 1936
27	22	1927	Dec. 22, 1926
34	13	1934	Jan. 13, 1934
38	08	1938	Jan. 8, 1938
47	04	1947	Jan. 4, 1947

to the data for Tel Aviv. Although the rainfall series for Haifa and Jerusalem (Table 1) are longer, Tel Aviv's data were selected as the dates of its peaks show a better agreement with the corresponding data for Sydney considered by Bowen.

In computing the accumulated daily rainfall, one writes, of course, the rainfall amounts for each date in one column. The reshuffling mentioned above consists of shifting the rows for the different years by a varying number of days. Thus, for instance, for one of the years shifted the rainfall amount for January 9 may fall in the column headed "January 1"; for the same year, the figure

for February 8 will appear in the column headed "January 31". As the Tel Aviv series comprises 27 years, dates between December 19 and January 14 inclusive were let appear once and only once in each case in the column marked "January 1". The years and dates were determined by reference to a table of randomly assorted digits published by Snedecor (1946, p. 12). In his table, two successive digits were thought to form a two-figured number. The tens and units of the figure representing the year (24-50) were obtained by reading Snedecor's rows from left to right; the days (19-31 and 01-14) were obtained by reading his columns downward, beginning with the left-hand column. For the years, it was sufficient to read the first nine rows; for the days, it was sufficient to read the first six (two-figured) columns. In this manner, a table of instructions, Table 2 above, for shifting the original data was derived, the two-figured numbers in Table 2 being quoted in the order as they appear for the first time in Snedecor's table.

The "accumulated daily rainfall amounts" resulting from shifting of the original data in accordance with Table 2 are shown in Figure 1 (b). It will be noted that the peaks of Figure 1 (a) for January 13, 20, 28, and February 2 have disappeared. New peaks have emerged for "January 7", "10", "29", and "February 3". Especially striking is the peak for "January 10". The "accumulated rainfall total" for this "date" exceeds by 20 per cent. the peak of Figure 1 (a) for January 13. In general, the character of the new fluctuation polygon is the same as that of the original.

Discussion of the Results

It is seen from the above that a random shifting of dates can lead to a polygon that simulates all essential characteristics of the original accumulated daily rainfall polygon. Although these results do not definitely disprove the hypothesis that the major peaks of Figure 1 (a) are produced by a genuine physical factor, they render such a hypothesis improbable. Indeed, they suggest that the peaks concerned have formed from a random accumulation of daily rainfall amounts. While this conclusion is primarily valid in respect of the Tel Aviv series, Bartels's (1948) work on the 97-year long rainfall series for Gütersloh, Westphalia, Germany, and the results of the present investigation would support the conjecture that the peaks for the stations considered by Bowen also are the results of random fluctuations.

A further shortcoming of Bowen's hypothesis is in that it only offers an explanation for the peaks but none for the "troughs", i.e. low rainfall amount totals. Thus, e.g., the data for Sydney (see Bowen's Fig. 1) indicate rather low values of accumulated daily rainfall on or about January 8, 16, 24, and February 4 which remain unaccounted for. The hypothesis of random accumulations could account for both the peaks and the troughs.

References

- BARTELS, J. (1948).—*Ann. Meteor.* **1**: 106-27.
 BOWEN, E. G. (1953).—*Aust. J. Phys.* **6**: 490-7.
 SNEDECOR, G. W. (1946).—"Statistical Methods." 4th Ed. (Iowa State College Press: Ames, Iowa.)

POLARIZATION OF BREMSSTRAHLUNG*

By E. G. MUIRHEAD† and K. B. MATHER‡

An attempt has been made to measure the polarization of radiation from an 11.5 MeV synchrotron. The method used was to load a photographic emulsion with heavy water, expose it with its plane normal to the beam, and measure the orientation and range of photoprotons from the ${}^2\text{H}(\gamma, p)n$ reaction. It has been confirmed experimentally (Wilkinson 1952) that protons from this reaction exhibit a \cos^2 distribution about the direction of the electric vector of plane polarized radiation, which means that the reaction may be used to reveal polarization.

An Ilford Nuclear Research Plate, 50 μ , C2 was loaded with D_2O by soaking for about 5 hr resulting in an absorption of 16.7 mg D_2O per cm^2 of plate. It was mounted in a watertight aluminium holder containing excess D_2O and exposed in the beam at 51 cm from the target, receiving 6 r from a 5 mil platinum target. A 20 cm lead collimator having a conical aperture prevented radiation from the doughnut reaching the plate and limited the exposure to an area of 5 cm diameter. Half an inch of "Bakelite" was interposed to reduce electron fogging. A similar exposure was made on H_2O loaded plates. Weight and thickness measurements were made on the plates giving water uptake, swelling and processing contraction factors, and permitting a calculation of the range-energy relation for the swollen emulsion (using data from Wilkins (1951)).

It was essential to know where the centre-line of the radiation cone intersected the plate. This was located by ionization chamber measurement at a distance of 4 m from the machine and fixed by a telescope sighting back to the target. The error was estimated as ≈ 2 mm at the plate. Scans were distributed over the plate to obtain data from a variety of planes of γ -emission. However, all the data used were obtained at an angular radius between 1 and 2° both to avoid the worst effects of an error in the centre-line and to concentrate results where polarization is likely to be most pronounced, $\theta \sim mc^2/E$.

The orientation ϕ of the projections of 1165 protons relative to the plane of emission of the γ -radiation is shown in Figure 1 broken down into three energy groups, *A*, *B*, and *C*. Background distributions (from H_2O loaded plates) were isotropic and should be subtracted from Figure 1 to the extent of ~ 20 per cent. in *A*, ~ 14 per cent. in *B*, and a few per cent. in *C*. § *A* is consistent with isotropy, a χ^2 test on *B* gave a probability $P = 0.3$ of getting this result if the distribution

* Manuscript received June 17, 1954.

† Physics Department, University of Melbourne.

‡ Australian Atomic Energy Commission, University of Melbourne.

§ Backgrounds were found to be essentially the same on unexposed regions of the same plates and presumably arose from old thorium α -tracks and single cosmic tracks.

is actually isotropic, and the probability of getting C as 0.01. Plotted with respect to an *arbitrary direction* on the plate the same set of tracks gave an isotropic distribution ($P=0.6$ for group C).

The 90° peak of Figure 1 C is consistent with a \sin^2 distribution and could be interpreted as a polarization perpendicular to the plane of emission. The 0° peak is more suggestive of a higher power of \cos^2 and its origin and shape, if taken seriously, are puzzling. Polarization can only give rise to a single peak, the electric vector being single-valued. However, the present work is not inconsistent with that of Phillips (1953) who reported peaks at 90° and $\sim 25^\circ$ in the total bremsstrahlung from a 20 MeV betatron.

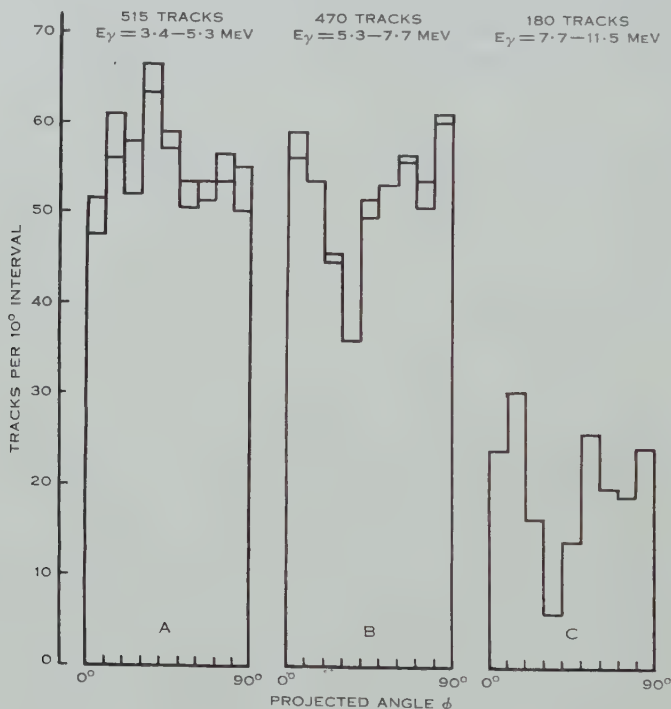


Fig. 1.—Projected angle, ϕ , distribution for three bremsstrahlung energy groups. ϕ is the angle between the projection of the track on the plane normal to the centre-line and the plane of emission of the γ -ray (p_0k plane of Gluckstern and Hull (1953)). The double ordinates in A and B indicate the number of tracks of uncertain origin.

A number of spurious effects have been considered and excluded as possible causes, e.g. multiple target traversals, a gross error in the centre-line determination, recoil protons, or deuterons knocked on by the photodisintegration neutrons in the loaded emulsion, etc. Contributions to the disintegration from effects other than the photoelectric (photomagnetic and tensor interaction) are of the order of a few per cent. and cannot appreciably alter the shape of the ϕ distribution.

The theory of polarization of bremsstrahlung (Gluckstern and Hull 1953) predicts a single peak at 90° which is expected to be more marked at energies corresponding to our cases *A* and *B*. However, before a quantitative comparison can be made, it is necessary to allow for effects of multiple scattering in the synchrotron target. Multiple scattering will reduce the degree of polarization predicted by Gluckstern and Hull towards isotropy. After due allowance is made for this effect it is difficult to reconcile the magnitude of the observed 90° peaks with theory. Work is continuing on this point and we are also repeating the experiment at 20 MeV.

Acknowledgment is made to Professor L. H. Martin for his interest in the progress of the experiment. Part of the scanning was done by Mrs. S. L. Ovenden.

References

- GLUCKSTERN, R. L., and HULL, M. H. (1953).—*Phys. Rev.* **90**: 1030.
PHILLIPS, K. (1953).—*Phil. Mag.* **44**: 169.
WILKINS, J. J. (1951).—A.E.R.E. Report G/R 664.
WILKINSON, D. H. (1952).—*Phil. Mag.* **43**: 659.

THE MULTIVARIATE t -DISTRIBUTION ASSOCIATED WITH A SET OF NORMAL SAMPLE DEVIATES

By E. A. CORNISH*

[Manuscript received July 29, 1954]

Summary

This paper gives a short account of the more important properties of the multivariate t -distribution, which arises in association with a set of normal sample deviates.

I. INTRODUCTION

The multivariate t -distribution described herein was encountered in the course of a recent investigation of the frequency distribution of the spectrographic error which occurs in the D.C. arc excitation of samples of soil and plant ash (Oertel and Cornish 1953). The original observations consisted of triplicate and sextuplicate determinations of copper, manganese, molybdenum, and tin in samples containing varying amounts of each of these elements, and it was found that the variance of measured line intensity was proportional to the square of the mean intensity. A logarithmic transformation stabilized the variance, which, for any individual element, could then be estimated from all the samples involving that element, and by subsequent standardization all the values of the new metric could be placed on a comparable basis. The standardized variates appeared to be normally distributed, and it became desirable, for various reasons, to test this point.

In providing tests of normality, two general lines of approach have been followed:

- (1) a normal distribution is fitted to the sample data, and the χ^2 test of goodness of fit is applied;
- (2) certain functions of the sample moments are calculated, and the significance of their departure from expectations based on the assumption of normality is examined,

but neither of these procedures, as ordinarily used, was suitable owing to the particular nature of the data. For this reason, and as the situation is likely to arise frequently in practice, other means were sought to make the necessary tests. A very limited number of exact tests is available (Fisher 1946), and reasonably accurate approximate tests have been devised by Pearson and Welch (1937) using as a base either a standard test of type (2) above or the work of Geary (1935, 1936). To meet the situation, other exact tests for small samples

* Division of Mathematical Statistics, C.S.I.R.O., Adelaide.

are required to supplement those given by Fisher, and consequently this particular t -distribution has been studied because it offered distinct possibilities in this connexion.

II. DISTRIBUTION FUNCTION OF NORMAL SAMPLE DEVIATES

Suppose x_1, x_2, \dots, x_n is a random sample of observations from a normal distribution specified by a mean ξ and variance σ^2 .

Let

$$\bar{x} = \sum_{i=1}^n x_i / n$$

be the sample mean, and

$$y_i = x_i - \bar{x}$$

be the deviates from the sample mean. The distribution function of one deviate is a classical result; Fisher (1920) gave the distribution of two deviates, and Irwin (1929) established the general non-singular distribution of $(n-1)$ deviates, which, without loss of generality, may be assumed to be the first $(n-1)$. In conformity with the notation used herein, the general distribution may be stated in the following manner:

If x_1, x_2, \dots, x_n are distributed in a multivariate normal distribution with variance-covariance matrix $\sigma^2 \mathbf{I}_n$,* then y_1, y_2, \dots, y_{n-1} and \bar{x} are distributed in a non-singular multivariate normal distribution with variance-covariance matrix

$$\mathbf{V} = \sigma^2 \begin{bmatrix} (n-1)/n & -1/n & \cdot & \cdot & \cdot & -1/n & 0 \\ -1/n & (n-1)/n & \cdot & \cdot & \cdot & -1/n & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ -1/n & -1/n & \cdot & \cdot & \cdot & (n-1)/n & 0 \\ 0 & 0 & \cdot & \cdot & \cdot & 0 & 1/n \end{bmatrix}$$

of order $n \times n$.

The deviates y_1, y_2, \dots, y_{n-1} are thus distributed independently of \bar{x} . Alternatively, y_1, y_2, \dots, y_n are distributed in a singular multivariate normal distribution with variance-covariance matrix

$$\sigma^2 \begin{bmatrix} (n-1)/n & -1/n & \cdot & \cdot & \cdot & -1/n \\ -1/n & (n-1)/n & \cdot & \cdot & \cdot & -1/n \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ -1/n & -1/n & \cdot & \cdot & \cdot & (n-1)/n \end{bmatrix}$$

of order $n \times n$ and rank $(n-1)$.

III. THE MULTIVARIATE t -DISTRIBUTION

Let \mathbf{B} denote the leading submatrix of order $(n-1) \times (n-1)$ in the matrix \mathbf{V} of Section II. If s^2 is an estimate of σ^2 , based on ν degrees of freedom and

* The symbol \mathbf{I} will designate the unit matrix, and the attached subscript will indicate its order.

distributed independently of y_1, y_2, \dots, y_{n-1} , the distribution function of the y_i and s may be written

$$\frac{|\mathbf{B}|^{-\frac{1}{2}}}{(2\pi)^{(n-1)/2}} \frac{v^{v/2}}{2^{(v-2)/2} \sigma^v \Gamma(v/2)} \int \dots \int e^{-\mathbf{y}'\mathbf{B}^{-1}\mathbf{y}/2} s^{v-1} e^{-vs^2/2\sigma^2} ds d\mathbf{y},^*$$

the multiple integral being taken over the region defined by the inequalities

$$-\infty \leq y_i \leq Y_i, \quad i=1, 2, \dots, n-1, \\ 0 \leq s \leq S.$$

\mathbf{B}^{-1} is the reciprocal matrix of \mathbf{B} , and \mathbf{y}' is the row vector $[y_1 y_2 \dots y_{n-1}]$.

Make the non-singular transformation indicated by the matrix equation

$$\mathbf{t} = \mathbf{Q}\mathbf{y},$$

where \mathbf{Q} is a diagonal matrix, whose diagonal elements are each equal to

$$\frac{1}{s} \sqrt{\frac{n}{n-1}}.$$

The jacobian is $|\mathbf{Q}|^{-1}$, and the distribution function becomes

$$\frac{v^{v/2}}{(2\pi)^{(n-1)/2} 2^{(v-2)/2} \sigma^v \Gamma(v/2)} \int \dots \int |\mathbf{Q}\mathbf{B}\mathbf{Q}'|^{-\frac{1}{2}} e^{-\mathbf{t}'(\mathbf{Q}\mathbf{B}\mathbf{Q}')^{-1}\mathbf{t}/2} e^{-vs^2/2\sigma^2} s^{v-1} dt ds,$$

in which the domain of integration is defined by

$$-\infty \leq t_i \leq T_i, \quad i=1, 2, \dots, n-1, \\ 0 \leq s \leq S.$$

Since

$$\mathbf{Q}\mathbf{B}\mathbf{Q}' = \frac{\sigma^2}{s^2} \begin{bmatrix} 1 & -1/(n-1) & \cdot & \cdot & \cdot & -1/(n-1) \\ -1/(n-1) & 1 & \cdot & \cdot & \cdot & -1/(n-1) \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ -1/(n-1) & -1/(n-1) & \cdot & \cdot & \cdot & 1 \end{bmatrix} \\ = \frac{\sigma^2}{s^2} \mathbf{R} \text{ say,}$$

it follows that

$$|\mathbf{Q}\mathbf{B}\mathbf{Q}'|^{-\frac{1}{2}} = \left(\frac{s}{\sigma}\right)^{n-1} |\mathbf{R}|^{-\frac{1}{2}},$$

and integrating for s from 0 to ∞ , the multivariate distribution function of t_1, t_2, \dots, t_{n-1} is

$$\frac{\Gamma(v+n-1)/2}{(\pi v)^{(n-1)/2} \Gamma(v/2)} |\mathbf{R}|^{-\frac{1}{2}} \int \dots \int (1 + \mathbf{t}'\mathbf{R}^{-1}\mathbf{t}/v)^{-(v+n-1)/2} d\mathbf{t}, \dots (1)$$

the integral being taken over the region defined by the inequalities

$$-\infty \leq t_i \leq T_i, \quad i=1, 2, \dots, n-1.$$

The limiting form of the distribution function (1), as $v \rightarrow \infty$, is that of a multivariate normal distribution with variance-covariance matrix \mathbf{R} .

* The notation $\Gamma m/2$ denotes $\Gamma(m/2)$ throughout.

IV. PROPERTIES OF THE DISTRIBUTION

(a) *Mean Values*

The vector of mean values, $E(\mathbf{t})$, is obviously null, and consequently the variance-covariance matrix is

$$\begin{aligned} E(\mathbf{t}\mathbf{t}') &= E\left(\frac{n}{s^2(n-1)}\mathbf{y}\mathbf{y}'\right) \\ &= E\left(\frac{1}{s^2}\right)E\left(\frac{n}{n-1}\mathbf{y}\mathbf{y}'\right) \\ &= \frac{\nu}{\sigma^2(\nu-2)} \cdot \sigma^2\mathbf{R} \\ &= \frac{\nu}{\nu-2}\mathbf{R}. \quad \dots\dots\dots (2) \end{aligned}$$

(b) *Distribution of Linear Functions*

Suppose

$$\mathbf{x} = \mathbf{H}\mathbf{t}$$

are any $p \leq (n-1)$ linearly independent linear functions of the t_i . The distribution of these functions can be derived directly but is more conveniently obtained from the corresponding result for normally distributed variates, namely, that if y_1, y_2, \dots, y_{n-1} are distributed in a multivariate normal distribution with variance-covariance matrix \mathbf{B} , then the variates

$$\mathbf{z} = \mathbf{H}\mathbf{y}$$

are distributed in a multivariate normal distribution with variance-covariance matrix \mathbf{HBH}' .

The distribution function of the z_i is thus

$$\frac{|\mathbf{HBH}'|^{-\frac{1}{2}}}{(2\pi)^{p/2}} \int \dots \int e^{-\mathbf{z}'(\mathbf{HBH}')^{-1}\mathbf{z}/2} d\mathbf{z}$$

over the region defined by the inequalities

$$z_i \leq Z_i, \quad i=1, 2, \dots, p,$$

and consequently, the distribution function of

$$\mathbf{x} = \mathbf{H}\mathbf{t} = \mathbf{H}\mathbf{Q}\mathbf{y} = \mathbf{Q}\mathbf{z}$$

and s is

$$\frac{\nu^{\frac{1}{2}}\nu}{(2\pi)^{p/2}2^{(\nu-2)/2}\sigma^\nu\Gamma_{\nu/2}} \int \dots \int |\mathbf{QHBH}'\mathbf{Q}'|^{-\frac{1}{2}} e^{-\mathbf{x}'(\mathbf{QHBH}'\mathbf{Q}')^{-1}\mathbf{x}/2} e^{-\nu s^2/2\sigma^2 s^{\nu-1}} ds d\mathbf{z}$$

over the region defined by

$$\begin{aligned} x_i &\leq X_i, & i=1, 2, \dots, p, \\ s &\leq S. \end{aligned}$$

Since

$$\mathbf{QHBH}'\mathbf{Q}' = \frac{\sigma^2}{s^2}\mathbf{HRH}',$$

integration for s from 0 to ∞ leaves the distribution of x_1, x_2, \dots, x_p in the form

$$\frac{\Gamma(\nu+p)/2 \mid \mathbf{H}\mathbf{R}\mathbf{H}' \mid^{-\frac{1}{2}}}{(\pi\nu)^{p/2}\Gamma\nu/2} \int \dots \int \{1 + \mathbf{x}'(\mathbf{H}\mathbf{R}\mathbf{H}')^{-1}\mathbf{x}/\nu\}^{-(\nu+p)/2} d\mathbf{x} \dots (3)$$

over the region defined by

$$x_i \leq X_i, \quad i=1, 2, \dots, p,$$

that is, a multivariate t -distribution of order p characterized by the matrix $(\mathbf{H}\mathbf{R}\mathbf{H}')^{-1}$.

(c) *Marginal Distribution of t_1, t_2, \dots, t_r*

The marginal distribution of $t_1, t_2, \dots, t_r, r < (n-1)$, follows from the previous result by taking

$$\mathbf{H} = \begin{bmatrix} \mathbf{I}_r & \cdot \\ \cdot & \cdot \end{bmatrix},$$

and is thus a multivariate t -distribution of order r , characterized by the matrix \mathbf{R}_1^{-1} , \mathbf{R}_1 being the leading submatrix of order $r \times r$ in \mathbf{R} . The variance-covariance matrix is $\{\nu/(\nu-2)\}\mathbf{R}_1$, and consequently, the variances and covariances of the marginal distribution are identical with their values in the original distribution. The limiting form of this distribution as $\nu \rightarrow \infty$ is the marginal multivariate normal distribution with variance-covariance matrix \mathbf{R}_1 .

(d) *Conditional Distribution of t_1, t_2, \dots, t_r*

To find the conditional distribution of t_1, t_2, \dots, t_r when t_{r+1}, \dots, t_{n-1} are given specified values, first partition the matrix \mathbf{R}^{-1} so that it takes the form

$$\begin{bmatrix} \mathbf{R}_1 & \mathbf{R}_3 \\ \mathbf{R}_3' & \mathbf{R}_2 \end{bmatrix}$$

where the submatrices \mathbf{R}_1 , \mathbf{R}_2 , and \mathbf{R}_3 respectively are of orders $r \times r$, $\{n-(r+1)\} \times \{n-(r+1)\}$, and $r \times \{n-(r+1)\}$ and the row vector \mathbf{t}' so that

$$\mathbf{t}' = [t_1 \dots t_r \mid t_{r+1} \dots t_{n-1}],$$

which may be written

$$[\mathbf{t}'_1 \quad \mathbf{t}'_2].$$

Since

$$\begin{bmatrix} \mathbf{I}_r & \cdot \\ -\mathbf{R}_3'\mathbf{R}_1^{-1} & \mathbf{I}_{n-(r+1)} \end{bmatrix} \begin{bmatrix} \mathbf{R}_1 & \mathbf{R}_3 \\ \mathbf{R}_3' & \mathbf{R}_2 \end{bmatrix} \begin{bmatrix} \mathbf{I}_r & -\mathbf{R}_1^{-1}\mathbf{R}_3 \\ \cdot & \mathbf{I}_{n-(r+1)} \end{bmatrix} = \begin{bmatrix} \mathbf{R}_1 & \cdot \\ \cdot & \mathbf{R}_2 - \mathbf{R}_3'\mathbf{R}_1^{-1}\mathbf{R}_3 \end{bmatrix}, \dots (4)$$

on taking reciprocals

$$(\mathbf{R}_2 - \mathbf{R}_3'\mathbf{R}_1^{-1}\mathbf{R}_3)^{-1} = \mathbf{R}_2,$$

where \mathbf{R}_2 is the submatrix in \mathbf{R} corresponding to \mathbf{R}_2 in \mathbf{R}^{-1} .

The marginal distribution of $t_{r+1}, t_{r+2}, \dots, t_{n-1}$ may thus be written

$$\frac{\Gamma\{\nu+n-(r+1)\}/2 \mid \mathbf{R}_2 - \mathbf{R}_3'\mathbf{R}_1^{-1}\mathbf{R}_3 \mid^{\frac{1}{2}}}{(\pi\nu)^{\{n-(r+1)\}/2}\Gamma\nu/2} \{1 + \mathbf{t}'_2(\mathbf{R}_2 - \mathbf{R}_3'\mathbf{R}_1^{-1}\mathbf{R}_3)\mathbf{t}_2/\nu\}^{-\{\nu+n-(r+1)\}/2} d\mathbf{t}_2.$$

Express the quadratic form $\mathbf{t}'\mathbf{R}^{-1}\mathbf{t}$ as

$$\mathbf{t}'\mathbf{R}^{-1}\mathbf{t}=\mathbf{t}_1'\mathbf{R}_1\mathbf{t}_1+2\mathbf{t}_2'\mathbf{R}_3'\mathbf{t}_1+\mathbf{t}_2'\mathbf{R}_3'\mathbf{R}_1^{-1}\mathbf{R}_3\mathbf{t}_2+\mathbf{t}_2'(\mathbf{R}_2-\mathbf{R}_3'\mathbf{R}_1^{-1}\mathbf{R}_3)\mathbf{t}_2,$$

and take as the fixed set of values for the variables in \mathbf{t}_2 , the elements of a vector \mathbf{a} . Using the determinantal relation found by taking determinants of both sides of (4), the conditional distribution function of \mathbf{t}_1 is

$$\frac{\Gamma(\nu+n-1)/2 \mid \mathbf{R}_1 \mid^{\frac{1}{2}}}{(\pi\nu)^{r/2}\Gamma\{\nu+n-(r+1)\}/2}\{1+\mathbf{a}'(\mathbf{R}_2-\mathbf{R}_3'\mathbf{R}_1^{-1}\mathbf{R}_3)\mathbf{a}/\nu\}^{\{\nu+n-(r+1)\}/2} \\ \times \int \dots \int \left\{1+\frac{(\mathbf{t}_1+\mathbf{R}_1^{-1}\mathbf{R}_3\mathbf{a})'\mathbf{R}_1(\mathbf{t}_1+\mathbf{R}_1^{-1}\mathbf{R}_3\mathbf{a})+\mathbf{a}'(\mathbf{R}_2-\mathbf{R}_3'\mathbf{R}_1^{-1}\mathbf{R}_3)\mathbf{a}}{\nu}\right\}^{-(\nu+n-1)/2} d\mathbf{t}_1, \\ \dots\dots\dots (5)$$

integration being taken over the domain specified by

$$t_i \leq T_i, \quad i=1,2,\dots,r.$$

The distribution function (5) is dependent upon the particular set of values chosen for the variables in the vector \mathbf{t}_2 , and its limiting form, as $\nu \rightarrow \infty$, is the conditional multivariate normal distribution of order r , with vector of means $-\mathbf{R}_1^{-1}\mathbf{R}_3\mathbf{a}$, and variance-covariance matrix \mathbf{R}_1^{-1} .

The mean value of \mathbf{t}_1 is

$$E(\mathbf{t}_1)=\frac{\Gamma(\nu+n-1)/2 \mid \mathbf{R}_1 \mid^{\frac{1}{2}}}{(\pi\nu)^{r/2}\Gamma\{\nu+n-(r+1)\}/2}\{1+\mathbf{a}'(\mathbf{R}_2-\mathbf{R}_3'\mathbf{R}_1^{-1}\mathbf{R}_3)\mathbf{a}/\nu\}^{\{\nu+n-(r+1)\}/2} \\ \times \int_{-\infty}^{\infty} \dots \int \mathbf{t}_1 \left\{1+\frac{(\mathbf{t}_1+\mathbf{R}_1^{-1}\mathbf{R}_3\mathbf{a})'\mathbf{R}_1(\mathbf{t}_1+\mathbf{R}_1^{-1}\mathbf{R}_3\mathbf{a})+\mathbf{a}'(\mathbf{R}_2-\mathbf{R}_3'\mathbf{R}_1^{-1}\mathbf{R}_3)\mathbf{a}}{\nu}\right\}^{-(\nu+n-1)/2} d\mathbf{t}_1.$$

To evaluate this integral, first make the change of variable

$$\mathbf{t}_1+\mathbf{R}_1^{-1}\mathbf{R}_3\mathbf{a}=\mathbf{z};$$

the jacobian is 1, and, omitting the constant factor, the integral becomes

$$\int_{-\infty}^{\infty} \dots \int (\mathbf{z}-\mathbf{R}_1^{-1}\mathbf{R}_3\mathbf{a}) \left\{1+\frac{\mathbf{z}'\mathbf{R}_1\mathbf{z}+\mathbf{a}'(\mathbf{R}_2-\mathbf{R}_3'\mathbf{R}_1^{-1}\mathbf{R}_3)\mathbf{a}}{\nu}\right\}^{-(\nu+n-1)/2} d\mathbf{z} \\ =-\mathbf{R}_1^{-1}\mathbf{R}_3\mathbf{a} \int_{-\infty}^{\infty} \dots \int \left\{1+\frac{\mathbf{z}'\mathbf{R}_1\mathbf{z}+\mathbf{a}'(\mathbf{R}_2-\mathbf{R}_3'\mathbf{R}_1^{-1}\mathbf{R}_3)\mathbf{a}}{\nu}\right\}^{-(\nu+n-1)/2} d\mathbf{z}.$$

Since \mathbf{R}_1 is a real, positive definite symmetric matrix, the quadratic form $\mathbf{z}'\mathbf{R}_1\mathbf{z}$ may be reduced to a sum of squares, after which the integration is easily performed, and finally yields

$$E(\mathbf{t}_1)=-\mathbf{R}_1^{-1}\mathbf{R}_3\mathbf{a}, \quad \dots\dots\dots (6)$$

which is a linear function of the elements of the vector \mathbf{a} . The regression of \mathbf{t}_1 on \mathbf{t}_2 thus exists and is linear.

The variance-covariance matrix is then

$$E(\mathbf{t}_1 + \mathbf{R}_1^{-1} \mathbf{R}_3 \mathbf{a})(\mathbf{t}_1 + \mathbf{R}_1^{-1} \mathbf{R}_3 \mathbf{a})' = \frac{\Gamma(\nu + n - 1)/2 \|\mathbf{R}_1\|^{1/2}}{(\pi \nu)^{r/2} \Gamma\{\nu + n - (r + 1)\}/2} \{1 + \mathbf{a}'(\mathbf{R}_2 - \mathbf{R}_3' \mathbf{R}_1^{-1} \mathbf{R}_3) \mathbf{a} / \nu\} \{\nu + n - (r + 1)\}^{1/2} \\ \times \int_{-\infty}^{\infty} \dots \int (\mathbf{t}_1 + \mathbf{R}_1^{-1} \mathbf{R}_3 \mathbf{a})(\mathbf{t}_1 + \mathbf{R}_1^{-1} \mathbf{R}_3 \mathbf{a})' \\ \left\{ 1 + \frac{(\mathbf{t}_1 + \mathbf{R}_1^{-1} \mathbf{R}_3 \mathbf{a})' \mathbf{R}_1 (\mathbf{t}_1 + \mathbf{R}_1^{-1} \mathbf{R}_3 \mathbf{a}) + \mathbf{a}'(\mathbf{R}_2 - \mathbf{R}_3' \mathbf{R}_1^{-1} \mathbf{R}_3) \mathbf{a}}{\nu} \right\}^{-(\nu + n - 1)/2} d\mathbf{t}_1,$$

and this integral is also readily evaluated after making the congruent transformation

$$\mathbf{t}_1 + \mathbf{R}_1^{-1} \mathbf{R}_3 \mathbf{a} = \mathbf{P} \mathbf{z}$$

such that $\mathbf{P}' \mathbf{R}_1 \mathbf{P} = \mathbf{I}_r$.

The variance-covariance matrix is

$$\frac{\nu + \mathbf{a}'(\mathbf{R}_2 - \mathbf{R}_3' \mathbf{R}_1^{-1} \mathbf{R}_3) \mathbf{a}}{\nu + n - (r + 3)} \mathbf{R}_1^{-1}, \dots \dots \dots (7)$$

which depends upon the particular values chosen for the variables in the vector \mathbf{t}_2 .

The limiting form of the matrix (7), as $\nu \rightarrow \infty$, is \mathbf{R}_1^{-1} , which is the variance-covariance matrix of the limiting conditional multivariate normal distribution, independent, as it should be, of the values of the fixed variates.

(c) *Distribution Function of $\mathbf{t}' \mathbf{R}^{-1} \mathbf{t}$*

(i) From the transformation of Section III

$$\mathbf{t} = \mathbf{Q} \mathbf{y},$$

and so

$$\mathbf{t}' \mathbf{R}^{-1} \mathbf{t} = \mathbf{y}' \mathbf{Q}' \mathbf{R}^{-1} \mathbf{Q} \mathbf{y} \\ = \sum_{i=1}^n y_i^2 / s^2.$$

Consequently

$$\mathbf{t}' \mathbf{R}^{-1} \mathbf{t} / (n - 1)$$

is distributed as e^{2z} with $(n - 1)$ and ν degrees of freedom.

(ii) The distribution of $\mathbf{t}' \mathbf{R}^{-1} \mathbf{t}$ may now be examined when the variables are subject to the linearly independent linear homogeneous conditions represented by the matrix equation

$$\mathbf{S} \mathbf{t} = \mathbf{0},$$

where \mathbf{S} is of order $p \times (n - 1)$ and rank $p < (n - 1)$.

Construct the matrix \mathbf{H} such that

$$\mathbf{H} \mathbf{R} \mathbf{H}' = \mathbf{I}_{n-p-1},$$

and

$$\mathbf{H} \mathbf{R} \mathbf{S}' = \mathbf{0},$$

and make the non-singular transformation

$$\mathbf{x} = \mathbf{P} \mathbf{t} = \begin{bmatrix} \mathbf{H} \\ \mathbf{S} \end{bmatrix} \mathbf{t}.$$

The distribution function of $\mathbf{t}'\mathbf{R}^{-1}\mathbf{t}$ thus becomes

$$\frac{|\mathbf{PRP}'|^{-\frac{1}{2}}\Gamma(\nu+n-1)/2}{(\pi\nu)^{(n-1)/2}\Gamma\nu/2}\int\ldots\int\{1+\mathbf{x}'(\mathbf{PRP}')^{-1}\mathbf{x}/\nu\}^{-(\nu+n-1)/2}d\mathbf{x},$$

where the domain of integration is defined by

$$\mathbf{x}'(\mathbf{PRP}')^{-1}\mathbf{x}\leq Q, \text{ say.}$$

Now impose the conditions

$$\mathbf{St}=\mathbf{0}.$$

This makes

$$x_{n-p}=x_{n-p+1}=\ldots=x_{n-1}=\mathbf{0}, \quad \ldots\ldots\ldots \quad (8)$$

and, using the results of Section IV (d), the conditional distribution of $x_1, x_2, \ldots, x_{n-p-1}$ when (8) holds is

$$\frac{\Gamma(\nu+n-1)/2}{(\pi\nu)^{(n-p-1)/2}\Gamma(\nu+p)/2}\int\ldots\int(1+\mathbf{x}'_1\mathbf{x}_1/\nu)^{-(\nu+n-1)/2}d\mathbf{x}_1,$$

where \mathbf{x}'_1 is the vector $[x_1, x_2, \ldots, x_{n-p-1}]$ and the integral is taken over the region defined by

$$\mathbf{x}'_1\mathbf{x}_1\leq Q.$$

A spherical polar transformation in $(n-p-1)$ dimensions, followed by a change of variable which makes the square of the radius vector equal to $\frac{\nu}{\nu+p}\chi^2$ reduces this integral to the form

$$\frac{\Gamma(\nu+n-1)/2}{(\nu+p)^{(n-p-1)/2}\Gamma(n-p-1)/2\Gamma(\nu+p)/2}\int_0^Q(\chi^2)^{(n-p-3)/2}\{1+\chi^2/(\nu+p)\}^{-(\nu+n-1)/2}d(\chi^2),$$

..... (9)

which is equivalent to Fisher's z -distribution with $(n-p-1)$ and $(\nu+p)$ degrees of freedom. The transfer of p degrees of freedom from one set of degrees of freedom to the other is a consequence of the fact that the conditional distribution is dependent upon the values of the fixed variates.

The distinction between the distribution (9) and that of $\mathbf{t}'\mathbf{R}^{-1}\mathbf{t}=\sum_{i=1}^ny_i^2/s^2$ when the y_i are subject to the restrictions $\mathbf{Sy}=\mathbf{0}$, should be noted. The latter distribution is, of course, Fisher's z -distribution with degrees of freedom $(n-p-1)$ and ν . Since s is essentially positive, either set of restrictive conditions then implies the other, and at first sight it might seem that the two distributions are identical.

(f) *Distribution Function of $\mathbf{t}'\mathbf{A}\mathbf{t}$*

The distribution function of the quadratic form $\mathbf{t}'\mathbf{A}\mathbf{t}$, of rank $r\leq(n-1)$, is given by

$$\frac{\Gamma(\nu+n-1)/2}{(\pi\nu)^{(n-1)/2}\Gamma\nu/2}\int\ldots\int(1+\mathbf{t}'\mathbf{R}^{-1}\mathbf{t}/\nu)^{-(\nu+n-1)/2}d\mathbf{t},$$

where the domain of integration is defined by

$$\mathbf{t}'\mathbf{A}\mathbf{t} \leq Q.$$

Make the congruent transformation

$$\mathbf{t} = \mathbf{H}\mathbf{x},$$

the matrix \mathbf{H} being chosen so that

$$\mathbf{H}\mathbf{R}^{-1}\mathbf{H}' = \mathbf{I}_{n-1},$$

and

$$\mathbf{H}\mathbf{A}\mathbf{H}' = \mathbf{\Lambda},$$

where $\mathbf{\Lambda}$ is a diagonal matrix whose diagonal elements are the roots of the equation

$$|\lambda\mathbf{R}^{-1} - \mathbf{A}| = 0,$$

or, alternatively, the latent roots of the matrix $\mathbf{R}\mathbf{A}$.

The jacobian is $|\mathbf{R}|^{\frac{1}{2}}$, so that the distribution function becomes

$$\frac{\Gamma(v+n-1)/2}{(\pi v)^{(n-1)/2} \Gamma v/2} \int \dots \int (1 + \mathbf{x}'\mathbf{x}/v)^{-(v+n-1)/2} d\mathbf{x},$$

the integral being taken over the region defined by

$$\mathbf{x}'\mathbf{\Lambda}\mathbf{x} \leq Q,$$

or

$$\sum_{i=1}^r \lambda_i x_i^2 \leq Q,$$

where $\lambda_1, \lambda_2, \dots, \lambda_r$ are the non-zero latent roots of $\mathbf{R}\mathbf{A}$.

After integrating for $x_{r+1}, x_{r+2}, \dots, x_{n-1}$, the distribution function becomes

$$\frac{\Gamma(v+r)/2}{(\pi v)^{r/2} \Gamma v/2} \int \dots \int (1 + \mathbf{x}'_1 \mathbf{x}_1/v)^{-(v+r)/2} d\mathbf{x}_1, \dots \quad (10)$$

where \mathbf{x}'_1 is the vector $[x_1 \ x_2 \ \dots \ x_r]$, and the domain of integration is defined by

$$\sum_{i=1}^r \lambda_i x_i^2 \geq Q.$$

Consequently, the necessary and sufficient condition that the distribution function (10) is equivalent to the z -distribution with degrees of freedom r and v , is that the non-zero latent roots of the matrix $\mathbf{R}\mathbf{A}$ are all equal to unity.

V. PARAMETRIC VALUES OF THE MULTIVARIATE DISTRIBUTION

In further discussion of the conditional distribution and of regression and correlation among the t -variates, consideration is given in particular to the case $r=1$, since this value is of the greatest importance in practice.

(a) Conditional Distribution

The determinant

$$|\mathbf{R}| = n^{n-2}/(n-1)^{n-1},$$

and, if R_{ij} denotes the co-factor of the (ij) th element in $|\mathbf{R}|$, then

$$R_{ii} = 2n^{n-3}/(n-1)^{n-2},$$

$$R_{ij} = n^{n-3}/(n-1)^{n-2}, \quad i \neq j,$$

and hence

$$\mathbf{R}^{-1} = \begin{bmatrix} 2(n-1)/n & (n-1)/n & \cdot & \cdot & \cdot & (n-1)/n \\ (n-1)/n & 2(n-1)/n & \cdot & \cdot & \cdot & (n-1)/n \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ (n-1)/n & (n-1)/n & \cdot & \cdot & \cdot & 2(n-1)/n \end{bmatrix}.$$

Taking $r=1$, the relation

$$\mathbf{R}^{-1} = \begin{bmatrix} \mathbf{R}_1 & \mathbf{R}_3 \\ \mathbf{R}'_3 & \mathbf{R}_2 \end{bmatrix}$$

gives

$$\mathbf{R}_1 = 2(n-1)/n,$$

and

$$\mathbf{R}_3 = [(n-1)/n \ (n-1)/n \ \cdot \ \cdot \ \cdot \ (n-1)/n],$$

a row vector of order $(n-2)$.

The conditional mean value of the variate t_1 is then

$$-\mathbf{R}_1^{-1} \mathbf{R}_3 \mathbf{a} = -\frac{1}{2} \sum_{j=2}^{n-1} a_j.$$

Moreover,

$$\mathbf{R}_2 - \mathbf{R}'_3 \mathbf{R}_1^{-1} \mathbf{R}_3 = \begin{bmatrix} 3(n-1)/2n & (n-1)/2n & \cdot & \cdot & \cdot & (n-1)/2n \\ (n-1)/2n & 3(n-1)/2n & \cdot & \cdot & \cdot & (n-1)/2n \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ (n-1)/2n & (n-1)/2n & \cdot & \cdot & \cdot & 3(n-1)/2n \end{bmatrix},$$

and hence

$$\nu + \mathbf{a}'(\mathbf{R}_2 - \mathbf{R}'_3 \mathbf{R}_1^{-1} \mathbf{R}_3) \mathbf{a} = \nu + \sum_j 3(n-1)a_j^2/2n + 2 \sum_{j < k} (n-1)a_j a_k/2n.$$

The conditional variance of the variate t_1 thus becomes

$$\frac{\nu + 3(n-1) \sum_j a_j^2/2n + (n-1) \sum_{j < k} a_j a_k/n}{\nu + n - 4} \frac{n}{2(n-1)},$$

and the average value of this quantity, for all possible values of a_2, a_3, \dots, a_{n-1} , is

$$\frac{\nu + \frac{3(n-1)}{2n}(n-2) \frac{\nu}{\nu-2} - \frac{n-1}{n} \frac{1}{2}(n-3)(n-2) \frac{\nu}{(\nu-2)(n-1)}}{\nu + n - 4} \frac{n}{2(n-1)} = \frac{\nu}{\nu-2} \frac{n}{2(n-1)}.$$

(b) Regression and Correlation

From (2), the variance-covariance matrix of the determining variates t_2, t_3, \dots, t_{n-1} is

$$\frac{\nu}{\nu-2} \mathbf{R}_2,$$

of which the determinant is

$$\frac{2n^{n-3}}{(n-1)^{n-2}} \left(\frac{\nu}{\nu-2} \right)^{n-2},$$

and, consequently, the reciprocal matrix

$$\left[\frac{\nu}{\nu-2} \mathbf{R}_2 \right]^{-1} = \begin{bmatrix} \frac{3(n-1)}{2n} & \frac{\nu-2}{\nu} & & & & & \frac{n-1}{2n} & \frac{\nu-2}{\nu} \\ \frac{n-1}{2n} & \frac{\nu-2}{\nu} & \frac{3(n-1)}{2n} & \frac{\nu-2}{\nu} & & & \frac{n-1}{2n} & \frac{\nu-2}{\nu} \\ & & & & & & & \\ & & & & & & & \\ \frac{n-1}{2n} & \frac{\nu-2}{\nu} & \frac{n-1}{2n} & \frac{\nu-2}{\nu} & & & \frac{3(n-1)}{2n} & \frac{\nu-2}{\nu} \end{bmatrix}.$$

Also from (2), the covariance of t_1 with each of t_2, t_3, \dots, t_{n-1} is

$$-\nu/(\nu-2)(n-1),$$

so that the vector of regression coefficients in the multiple regression of t_1 on t_2, t_3, \dots, t_{n-1} is

$$\left[\frac{\nu}{\nu-2} \mathbf{R}_2 \right]^{-1} \begin{bmatrix} \frac{-\nu}{(\nu-2)(n-1)} \\ \frac{-\nu}{(\nu-2)(n-1)} \\ \vdots \\ \frac{-\nu}{(\nu-2)(n-1)} \end{bmatrix} = \begin{bmatrix} -\frac{1}{2} \\ -\frac{1}{2} \\ \vdots \\ -\frac{1}{2} \end{bmatrix},$$

in agreement with the coefficients of the linear function for the conditional mean value of t_1 .

The residual variance of t_1 with respect to t_2, t_3, \dots, t_{n-1} is

$$\frac{\left| \frac{\nu}{\nu-2} \mathbf{R} \right|}{\left| \frac{\nu}{\nu-2} \mathbf{R}_2 \right|} = \frac{\nu}{\nu-2} \frac{n}{2(n-1)},$$

independent of the values of the determining variates, and equal to the average conditional variance given above.

The square of the multiple correlation of t_1 with t_2, t_3, \dots, t_{n-1} is

$$1 - \frac{\left| \frac{\nu}{\nu-2} \mathbf{R} \right|}{\frac{\nu}{\nu-2} \left| \frac{\nu}{\nu-2} \mathbf{R}_2 \right|} = (n-2)/2(n-1),$$

and the partial correlation of t_1 with t_j ($j=2,3,\dots,n-1$) is

$$\frac{-\left(\frac{v}{v-2}\right)^{n-2} R_{1j}}{\sqrt{\left(\frac{v}{v-2}\right)^{n-2} R_{11} \left(\frac{v}{v-2}\right)^{n-2} R_{jj}}} = -\frac{1}{2}.$$

VI. REFERENCES

- FISHER, R. A. (1920).—*Mon. Not. R. Astr. Soc.* **80**: 758.
 FISHER, R. A. (1946).—"Statistical Methods for Research Workers." 10th Ed. (Oliver and Boyd: Edinburgh.)
 GEARY, R. C. (1935).—*Biometrika* **27**: 310.
 GEARY, R. C. (1936).—*Biometrika* **28**: 295.
 IRWIN, J. O. (1929).—*J. R. Statist. Soc.* **92**: 580.
 OERTEL, A. C., and CORNISH, E. A. (1953).—*Aust. J. Appl. Sci.* **4**: 489.
 PEARSON, E. S., and WELCH, B. L. (1937).—*J. R. Statist. Soc. Suppl.* **4**: 94.

THE ANGULAR DISTRIBUTION OF HIGH ENERGY ELECTRONS IN AIR SHOWERS

I. LANDAU APPROXIMATION

By M. H. KALOS* and J. M. BLATT†

[Manuscript received May 31, 1954]

Summary

The angular distribution of high energy ($>5 \times 10^8$ eV) electrons in air showers is calculated on a track length basis, using approximation A of Rossi and Greisen (1941) (no ionization loss) and the Landau (1940) multiple scattering approximation. We start with a discussion of the approximations used and an estimate of their validity. The basic equations are written down; qualitative results and very rough solutions are developed and applied to the Furry model of a cascade. For the Furry cascade the qualitative arguments lead directly to an Ansatz which yields an exact solution. In the actual cascade the corresponding Ansatz does not yield an exact solution. We then perform an iteration, employing a general method of Friedman. The final (iterated) solution is compared with the exact (in the Landau approximation) solution by means of their moments, and appears to be within 10 per cent. of the correct solution for $E\theta/E_s < 1$. Our solution compares well with earlier work on this problem. Appendices contain a short derivation of the angular moments, a general inversion formula for going from the distribution-in-projected-angle to the distribution-in-angle-with-the-shower-axis, and a derivation of the Friedman variation principle in vector space terminology.

I. INTRODUCTION

The interactions of extremely high energy cosmic rays can usually be observed only from the cascades they produce in the atmosphere (air showers). A theory of the cascade process including the angular and lateral distribution of the particles is necessary to interpret the observations and infer the nature of the primary events. For example, the knowledge of the expected lateral structure of the shower of electrons and photons produced by a single source will help determine whether an observed shower comes from one particle or from a number of lower energy sources. In this paper we discuss a rather limited aspect of the more general problem, namely, the angular distribution of monoenergetic electrons in a single-source shower.

Except for some work by Molière (1946) and by Belenky (1944), calculations on the angular distribution (Roberg and Nordheim 1949; Eyges and Fernbach 1951; Green and Messel 1952) have centred on the moments of this distribution function. This approach is simpler than a direct attack on the distribution function. Many functions can be determined over much of their range from

* University of Illinois, Urbana, Ill., U.S.A.; present address: Laboratory for Nuclear Studies, Cornell University, Ithaca, N.Y., U.S.A.

† School of Physics, University of Sydney.

the sequence of moments. Unfortunately, it is quite difficult to find the behaviour near zero in this way. Although there exist some mathematical theorems about the uniqueness of the moment inversion, these theorems are of little practical help.

The work described here was undertaken in an attempt to calculate the so-called "track length" values of the angular distribution with particular emphasis on the behaviour of the function for small argument. The track length distribution is obtained by integrating over the whole depth of the shower; it naturally contains less information than the original function with depth dependence included. It is considerably easier to work with, however, and can be used to analyse experiments in which all electrons above a given energy are detected. In such a situation, the distribution-in-angle at any one depth t is determined primarily from that part of the energy range which gives the largest number of particles at that depth. The distribution of these electrons is very close to the track length distribution for that same energy.*

We shall use the set of approximations which Rossi and Greisen (1941) call "Approximation A"; that is, ionization loss, Compton collisions, and knock-on collisions are ignored, and the cross sections for pair production and bremsstrahlung are replaced by their extreme relativistic values. This set of approximations is reasonable for electrons of energy much higher than the critical energy (much higher than 100 MeV in air). Furthermore, to simplify the mathematics, we shall use, instead of the actual extreme relativistic approximations for the cross sections, some mathematically more tractable expressions, namely,

$$\sigma_{\text{pair}}(E, E') = 7/(9E), \quad \sigma_{\text{brems}}(E, E') = (E - E')^{-1}. \quad \dots \quad (1.1)$$

Here E is the energy of the initial particle or photon and E' is the energy of a secondary electron. In the case of bremsstrahlung, the secondary photon has energy $(E - E')$, in the case of pair creation the second member of the pair has this energy. The cross sections (1.1) are called "super-simplified cross sections" by Friedman (1949); the cross sections differ from the correct ones by about 30 per cent. at most (see Rossi and Greisen 1941, Figs. 8 and 10), but of course the error in the resulting angular distribution function is smaller than this. The choice of supersimplified rather than conventional cross sections leads to errors of less than 10 per cent. in the region $E\theta/E_c < 1$.

As far as the angular spread of the shower is concerned, we shall ignore all spreading processes other than Coulomb scattering of electrons by the nuclei of the air. The most important processes neglected are the angular deviations in bremsstrahlung and pair creation, and the angular deviation due to the action

* The use of track length quantities has been questioned by Green and Messel (1952). In our opinion the approximation suggested above, when applied to air showers, is at least as good as some of the basic approximations (neglect of the angular spread in bremsstrahlung and pair production, neglect of the effect of the magnetic field of the Earth) made in all calculations so far, including the calculations of Green and Messel. Furthermore, it is rather easy to find approximation methods for taking into account the deviations from the track length distribution once the track length distribution itself is known.

of the magnetic field of the Earth.* All these processes have the same dependence on energy as the Coulomb scattering (E^{-1}). Their importance can be estimated by considering the mean square angles of deviation per radiation length :

$$\text{Multiple Coulomb scattering : } \langle \theta^2 \rangle_{\text{ave}} = (E_s/E)^2 \quad E_s = 21 \text{ MeV, } \dots (1.2)$$

$$\text{Bremsstrahlung : } \langle \theta^2 \rangle_{\text{ave}} \sim [(mc^2/E) \ln (E/mc^2)]^2, \dots (1.3)$$

$$\text{Pair production : } \langle \theta^2 \rangle_{\text{ave}} \sim [(mc^2/h\nu) \ln (h\nu/mc^2)]^2. \dots (1.4)$$

The effect of the magnetic field of the Earth depends upon the direction of the shower axis and upon the air pressure, being nearly zero if the shower axis is parallel to the magnetic field lines. However, for showers with directions appreciably different from that of the magnetic field of the Earth, $\langle \theta^2 \rangle_{\text{ave}}$ is of the same general magnitude as the $\langle \theta^2 \rangle_{\text{ave}}$ for multiple Coulomb scattering, especially so at mountain altitudes where the lower air pressure implies a longer actual path for a given path length in radiation lengths. The factor $\ln(E/mc^2)$ in the bremsstrahlung and pair production $\langle \theta^2 \rangle_{\text{ave}}$ becomes larger as the energies increase. However, in practice we are not interested in energies larger than 10^{10} eV, and for such energies $\langle \theta^2 \rangle_{\text{ave}}$ for the radiative processes is smaller than $\langle \theta^2 \rangle_{\text{ave}}$ for multiple Coulomb scattering by a factor larger than 16. Thus the neglect of the angular deviations in radiative processes should not lead to errors larger than 10 per cent. in the final angular distribution function. The neglect of the angular deviations due to the magnetic field of the Earth leads to errors larger than this unless the axis of the shower is substantially parallel to the direction of the magnetic field of the Earth. None of the work published so far has taken any of these processes into account.

Besides ignoring these other spreading processes, we shall use the Landau (1940) multiple scattering approximation for the Coulomb scattering of electrons. This approximation has been criticized lately by Green and Messel (1952), who point out that it leads to serious errors in the higher moments of the angular distribution function. Their calculations show that the Landau approximation leads to serious errors in the distribution function for values of $u = E\theta/E_s > 1$. Whether or not the Landau approximation also leads to serious errors in the distribution function for *small* values of u cannot be decided by a comparison of moments.

However, the qualitative arguments given by Rossi (1952), as well as the exact calculations of Snyder and Scott (1949) on the simpler diffusion problem without cascade multiplication, indicate strongly that the Landau approximation is in fact insufficient even for small values of $E\theta/E_s$. The angular distribution found by Snyder and Scott behaves roughly like a Gaussian curve for small angles, but the width of the Gaussian is not given correctly by the Landau approximation unless this width is larger than the maximum angle for single Coulomb scattering (Rossi's criterion of validity, eqn. (14), p. 72). For a

* The importance of this latter effect has been pointed out recently by Cocconi (1954).

typical path length of about $\frac{1}{2}$ radiation unit per electron in the shower* this criterion is badly violated. We therefore feel that the Landau approximation cannot be trusted even for small distances and angles.

There is some point in making calculations with the Landau approximation, nevertheless. First, all of the previous work not concerned only with moments has used this approximation, and hence a comparison is in order. Second, once it is realized that the main error of the Landau approximation for small angles arises from the error in the width of the Gaussian (for no cascade), a simple method of correction presents itself fairly obviously. It consists in altering the value of the energy E_s in such a way that the width of the Landau Gaussian curve agrees approximately with the width of the Scott and Snyder Gaussian for depths of the order of 0.1 – 0.5 radiation lengths. The efficacy of this method is now being tested, and will be the subject of a later publication. It should be emphasized that this method of correction works only for values of $E\theta/E_s \ll 1$. For larger angles, the moments computed by Green and Messel (1952) are sufficient to define the behaviour of the distribution function to sufficient accuracy.

The general approach of this paper is adapted to the study of air showers, and several of the approximations made here would not be applicable to other types of showers. We have tried to obtain a distribution function which is within 10 per cent. of the correct value for all angles which are experimentally measurable, and for all energies larger than 5×10^8 eV (at lower energies approximation A can not be trusted to this accuracy).

Section II of this paper gives the mathematical equations to be solved. In Section III we give some qualitative arguments; these arguments lead to a very rough guess about the behaviour of the distribution function. In Section IV these arguments are applied to the Furry model of a shower, and for this model it is shown that they lead naturally to an exact solution. Since the Furry cascade does not differ from the true cascade in features which are likely to have a strong influence on the angular distribution function, we can have some confidence that the qualitative arguments of Section III apply also to the actual shower. One result of Section IV is that the Furry cascade does not lead to any singularity in the distribution function $f(u)$ at $u=0$. It is extremely likely that there is also no singularity of $f(u)$ for the actual cascade. In Section V we apply the same qualitative arguments to obtain an approximate distribution function for the actual shower. Unlike the Furry cascade, this is not an exact solution. We obtain our final result in Section VI by using the approximate solution of Section V as a trial function in a variationally correct iteration procedure. The accuracy of the iterated function is tested in two ways: by comparison with the exact (in Landau approximation) moments, and by comparison with the results of Belenky (1944) and of Molière (1946). Both comparisons show that the iterated function is within 10 per cent. of the exact

* This is the mean free path against bremsstrahlung collisions in which the electron loses more than 10 per cent. of its energy. The mean free path for photons is not relevant here because photons are not scattered at all under our approximations.

function (in the Landau approximation) for values of $u < 1$. Appendix I contains calculations of the moments of the distribution function; Appendix II gives an explicit relationship between the distribution-in-projected-angle θ (as found in this paper) and the distribution-in-angle Θ between the direction of motion of the shower particle and the shower axis; Appendix III contains a general statement of the variational method used.

We realize that experimentalists are not primarily interested in the track length angular distribution of monoenergetic electrons, but would like to know the theoretically predicted lateral distribution at a given depth t from the origin of the shower, integrated over all electron energies. Work on that problem is in progress.

II. DIFFUSION EQUATIONS FOR THE ANGULAR DISTRIBUTION

Let $\varphi_\pi(E_0, E, t, \theta) dE d\theta$ be the average number of electrons in the energy interval $E, E + dE$ at depth t making a projected angle with the shower axis (direction of initial electron) in the range $\theta, \theta + d\theta$; the initial electron had energy $E = E_0$ at the point $t = 0$. Let $\varphi_\gamma(E_0, E, t, \theta)$ be the corresponding average distribution function for the photons. We introduce integral operators A, B, C as follows: let $f(E)$ be an arbitrary function of E , other variables being suppressed for the moment; then

$$Af = \lim_{\varepsilon \rightarrow 0} \left[- \int_{E+\varepsilon}^{\infty} f(E') \sigma_{\text{brem}}(E', E) dE' + f(E) \int_0^{E-\varepsilon} \sigma_{\text{brem}}(E, E') dE' \right], \quad \dots (2.1)$$

$$Bf = 2 \int_E^{\infty} f(E') \sigma_{\text{pair}}(E', E) dE', \quad \dots (2.2)$$

$$Cf = \int_E^{\infty} f(E') \sigma_{\text{brem}}(E', E' - E) dE'. \quad \dots (2.3)$$

The limiting process is necessary because the bremsstrahlung cross section contains the infra-red catastrophe. The signs are chosen in such a way that the quantities $A(s)$, $B(s)$, and $C(s)$ in Rossi and Greisen (1941) are the Mellin transform images of the operators A , B , and C respectively. We also use the notation σ_0 for the total pair production cross section; in approximation (1.1) $\sigma_0 = 7/9$. Finally $E_s = 21$ MeV is the characteristic energy for multiple Coulomb scattering. In what follows all angles are projected angles, that is, the whole shower is projected onto a plane containing the shower axis. This corresponds experimentally to taking measurements of angles in a cloud chamber without stereoscopic photographs. Furthermore, all angles are assumed to be much less than 1 radian.*

* This approximation is excellent for high energy electrons ($E \gg E_s$) and it has nothing to do with the question whether the moments of the angular distribution function determine the function for "large" or "small" angles. The characteristic variable in the angular distribution function is not θ itself, but the combination $E\theta/E_s$. For energies $E \gg E_s$ it is quite possible that $E\theta/E_s > 1$ and yet $\theta \ll 1$. Approximation A, which neglects ionization loss, is valid only when E exceeds the critical energy, which is about 4 times E_s in air. Hence the condition $E \gg E_s$ is automatically satisfied under the assumptions of our calculation.

In this paper we shall work with the Landau approximation throughout; the limitations of this approximation will be discussed in a later publication. The Landau (1940) diffusion equations are:

$$\partial\varphi_{\pi}/\partial t = -A\varphi_{\pi} + B\varphi_{\gamma} + (E_s/2E)^2\partial^2\varphi_{\pi}/\partial\theta^2 + \delta(E_0 - E)\delta(t)\delta(\theta), \quad \dots \quad (2.4)$$

$$\partial\varphi_{\gamma}/\partial t = C\varphi_{\pi} - \sigma_0\varphi_{\gamma}. \quad \dots \quad (2.5)$$

The delta function in (2.4) represents a single incident electron at $t=0$ with energy $E=E_0$ and direction $\theta=0$. Note that these equations are invariant under translation in t even in the case of an inhomogeneous medium such as the air. The changing density of the air has no effect on the angular distribution, although it does affect the lateral distribution.*

We now eliminate the photon distribution φ_{γ} from (2.4) by using (2.5). We get from (2.5):

$$\varphi_{\gamma} = (\partial/\partial t + \sigma_0)^{-1}C\varphi_{\pi}, \quad \dots \quad (2.6)$$

$$(d/dt + \sigma_0)^{-1}f(t) = \int_0^t f(t') \exp[\sigma_0(t' - t)] dt'. \quad \dots \quad (2.7)$$

We then substitute (2.6) into (2.4) and define the operator L by

$$L = \partial/\partial t + A - B(\partial/\partial t + \sigma_0)^{-1}C \quad \dots \quad (2.8)$$

to get an equation containing φ_{π} only

$$L\varphi_{\pi} - (E_s/2E)^2\partial^2\varphi_{\pi}/\partial\theta^2 = \delta(E_0 - E)\delta(t)\delta(\theta). \quad \dots \quad (2.9)$$

This is the fundamental diffusion equation for the Landau theory. From this equation others can be derived. A particularly useful integral relation is suggested by a moment recursion relation found by Nordheim (1952). Let us use the Rossi and Greisen notation $\pi(E_0, E, t)$ for the average number of electrons irrespective of angle, that is,

$$\pi(E_0, E, t) = \int_{-\infty}^{\infty} \varphi_{\pi}(E_0, E, t, \theta) d\theta. \quad \dots \quad (2.10)$$

The integration extends over an infinite range as a result of the assumption of small angles throughout the calculation. Let us denote by $\pi_n(E_0, E, t)$ the n th moment of the function φ_{π} , that is,

$$\pi_n(E_0, E, t) = \int_{-\infty}^{\infty} \theta^n \varphi_{\pi}(E_0, E, t, \theta) d\theta. \quad \dots \quad (2.11)$$

Clearly $\pi_0 = \pi$. The Nordheim moment recursion relation then reads

$$\pi_n(E_0, E, t) = n(n-1) \int_0^t dt' \int_E^{E_0} dE' (E_s/2E')^2 \pi(E_0, E', t') \times \pi_{n-2}(E', E, t-t'). \quad \dots \quad (2.12)$$

* The magnitude of the effect on the *lateral* distribution depends upon how one interprets the calculations for a homogeneous atmosphere when applying them to the actual inhomogeneous atmosphere. The usual method consists in measuring all distances, radial as well as vertical, in radiation units. The effect is then of the order of 10-20 per cent. at mountain altitudes. By measuring vertical distances in radiation units but radial distances in centimetres, the effect of the changing density of the atmosphere is greatly magnified. As Messel and Green (1952) point out, this accounts for the 5000 per cent. correction found by them.

Consider the integral equation

$$\varphi_{\pi}(E_0, E, t, \theta) = \frac{\partial^2}{\partial \theta^2} \int_0^t dt' \int_E^{E_0} dE' (E_s/2E')^2 \pi(E_0, E', t') \varphi_{\pi}(E', E, t-t', \theta) + \delta(\theta) \pi(E_0, E, t). \quad (2.13)$$

By multiplying both sides of (2.13) by θ^n and integrating over θ , it is seen that (2.13) is consistent with (2.12). Equation (2.13) can also be proved directly from the Landau equation (2.9). The proof is rather lengthy and the reader is referred to a thesis by one of us (Kalos 1952).

We shall be working mostly with track length quantities. We define the track length $v(E_0, E, \theta)$ by

$$v(E_0, E, \theta) = \int_0^{\infty} \varphi_{\pi}(E_0, E, t, \theta) dt. \quad (2.14)$$

The longitudinal track length is

$$z_{\pi}(E_0, E) = \int_{-\infty}^{\infty} v(E_0, E, \theta) d\theta = \int_0^{\infty} \pi(E_0, E, t) dt. \quad (2.15)$$

To the extent that the main contribution to the integral over the depth t comes from a narrow region of t , namely, that region over which electrons of energy E have their maximum number, the track length v is close to the value of $\varphi_{\pi}(E_0, E, t, \theta)$ at $t = t_{\max}$. The advantage of the track length is of course that it is much easier to work with.

By integrating over t from 0 to infinity, we get from (2.8)

$$Fv - (E_s/2E)^2 \partial^2 v / \partial \theta^2 = \delta(\theta) \delta(E_0 - E), \quad (2.16)$$

where the operator F is the track length analogue of the operator L , (2.8), and is defined by

$$F = A - BC / \sigma_0. \quad (2.17)$$

A similar procedure applied to the integral equation (2.13) gives

$$v(E_0, E, \theta) = \frac{\partial^2}{\partial \theta^2} \int_E^{E_0} dE' \left(\frac{E_s}{2E'} \right)^2 z_{\pi}(E_0, E') v(E', E, \theta) + \delta(\theta) z_{\pi}(E_0, E). \quad (2.18)$$

III. QUALITATIVE ARGUMENTS

Rather than plunge directly into the mathematics of finding solutions to these equations, let us first give some qualitative arguments to establish the general, rough features of the distribution functions we are looking for. We shall be particularly interested in the behaviour of $v(E_0, E, \theta)$ for two extreme cases: $E \ll E_0$, and E very close to E_0 .

It is clear physically that the nature and energy of the initial particle which started the cascade cannot have a great influence on the angular structure of the shower at energies $E \ll E_0$, even though the number of particles of this energy depends very much on E_0 . The high energy particles all stay close to the core of the shower, and the angular (as well as lateral) deviations observed for particles of energy E arise during the last few radiation lengths, that is, during the period

when the particle or its ancestors had energies not tremendously much larger than E . If $E \ll E_0$, this means that the observed deflexions arise from particles all of which had energies $E' \ll E_0$, and hence the particular value of E_0 cannot matter for the angular structure of the shower. Once this point is established, dimensional arguments applied to equation (2.16) show immediately that the angle θ can enter only in the combination

$$u = E\theta/E_s. \quad \dots\dots\dots (3.1)$$

The angular structure of the shower is defined essentially by the ratio v/z_π . Guided by the arguments above, we define the angular structure function $f(E_0, E, \theta)$ of the shower by

$$f(E_0, E, \theta) = (E_s/E)(v/z_\pi). \quad \dots\dots\dots (3.2)$$

We then expect,

$$f = f(u) \text{ only, } \int_{-\infty}^{\infty} f(u) du = 1, \quad \text{for } E \ll E_0. \quad \dots (3.3)$$

This structure function $f(u)$ is the information of primary interest. Its moments, in the Landau approximation, are given in Appendix I as well as in the paper by Eyges and Fernbach (1951). Unfortunately, it turns out that an understanding of $v(E_0, E, \theta)$ for values of E close to E_0 is necessary in order to get useful results for $f(u)$ when E is much less than E_0 .

In order to gain such understanding, let us for the moment ignore the infra-red divergence in the bremsstrahlung cross section; we shall assume that the total bremsstrahlung cross section is finite. It is then possible to expand according to successive collisions,* that is, we shall group the particles according to the number of radiative collisions which have occurred in their ancestry. The "zero group" or "end group" consists of the initial particle. This initial particle is multiply scattered, and its probability of surviving against radiative (bremsstrahlung) collisions is $\exp(-\sigma_1 t)$ where t is the thickness of matter traversed, and σ_1 is the (assumed finite) total bremsstrahlung cross section. The angular distribution for multiple scattering without cascade multiplication, in the Landau approximation, was given by Fermi (as quoted in Rossi and Greisen (1941)). It is

$$G(E, t, \theta) = (E/E_s)(\pi t)^{-\frac{1}{2}} \exp[-t^{-1}(E\theta/E_s)^2]. \quad \dots\dots (3.4)$$

Thus the distribution function for the end group of particles, that is, for the initial particle, is

$$\varphi_\pi^{(0)}(E_0, E, t, \theta) = G(E_0, t, \theta) e^{-\sigma_1 t \delta(E_0 - E)}, \quad \dots\dots\dots (3.5)$$

and the corresponding track length distribution function is

$$v^{(0)}(E_0, E, \theta) = \delta(E_0 - E) \int_0^\infty G(E_0, t, \theta) \exp(-\sigma_1 t) dt. \quad \dots (3.6)$$

* A much more complicated expansion in successive collisions is possible also with an infinite bremsstrahlung total cross section; this was given by H. J. Bhabha and W. Heitler in their fundamental paper on shower theory. In the Bhabha-Heitler expansion, the particles are grouped into generations according to the number of photons in their ancestry. The number of bremsstrahlung events is not counted, indeed that number is infinite.

The integration in (3.6) can be performed analytically by using an integral representation for Bessel functions of imaginary arguments given by Watson (1948, p. 183). The Bessel function involved is $K_{-\frac{1}{2}}$ which is expressible in terms of elementary functions. The result is

$$v^{(0)}(E_0, E, \theta) = (v/\sigma_1) \exp(-2v|\theta|) \delta(E_0 - E),$$

$$v = (E_0/E_s)(\sigma_1)^{\frac{1}{2}}. \quad \dots\dots\dots (3.7)$$

It is worth remarking that the discontinuity in the derivative of this function at $\theta=0$ corresponds to a (weak) infinity of the corresponding distribution in the angle Θ between the actual (not projected) motion of the particles and the shower axis. Indeed, the corresponding integration leads to the Bessel function K_0 which has a logarithmic infinity at the origin. The same result follows also from the inversion formula discussed in Appendix II, which allows one to go directly from the distribution in the projected angle to the distribution in the actual angle $\Theta = (\theta_x^2 + \theta_y^2)^{\frac{1}{2}}$.

The next generation has an angular distribution which is harder to compute analytically, and we shall not do so here. If for some reason the behaviour of v for $E=E_0$ (or for E very close to E_0 in the true cascade) is sufficient, expression (3.7) can be used.

Unfortunately these considerations cannot be applied directly to the actual shower, since the bremsstrahlung cross section σ_1 diverges. This has a considerable influence on the detailed form of the distribution function for E near E_0 . Let us consider as the end group of the actual shower the initial electron which may have undergone any number of bremsstrahlung collisions, but has no photon in its ancestry. Then a good first approximation for the longitudinal distribution function of this end group is*

$$\pi^{(0)}(E_0, E, t) = \frac{[\ln(E_0/E)]^{(t/\ln 2) - 1}}{E_0[(t/\ln 2) - 1]!} \quad \dots\dots\dots (3.8)$$

This should be compared with the expression

$$\pi^{(0)}(E_0, E, t) = \delta(E_0 - E) \exp(-\sigma_1 t) \quad \dots\dots\dots (3.9)$$

for the end group of a hypothetical shower with finite total bremsstrahlung cross section σ_1 . Of course, the comparison is not really warranted because the expression "end group" denotes two different things in these two cases. But for our purposes, (3.8) is adequate for energies E close to E_0 in the actual shower, (3.9) is adequate in the hypothetical shower, and in this sense the two are comparable. We see that, at least for reasonably small t , (3.8) is very sharply peaked near $E=E_0$, but that peak is not nearly as strong as the delta function peak in (3.9).

In the angular problem, a reasonably good approximation for the end group can be obtained by multiplying (3.8) by the Fermi function $G(E_0, t, \theta)$. Unfortunately, the track length $v^{(0)}$ which then replaces (3.6) cannot be found in simple closed form. Since we need a simple closed form for our later work, we are forced to use a very much rougher approximation.

* See Rossi (1952, p. 244). This formula was first found by Bethe and Heitler and is sometimes called the range straggling formula for bremsstrahlung.

We shall employ the expression (3.9) for the end group of the actual shower, with the constant σ_1 adjusted in some "best" way. The adjustment follows the suggestion of Friedman (1949). It is well known that in the actual shower the longitudinal track length $z_\pi(E_0, E)$ is given by

$$z_\pi(E_0, E) = \alpha E_0/E^2, \quad \text{for } E \ll E_0, \dots\dots\dots (3.10)$$

where $\alpha = 0.4368$ for "conventional" cross section and $\alpha = 0.4662$ for "super-simplified" cross section. The longitudinal track length associated with (3.9) is

$$z_\pi^{(0)}(E_0, E) = (\sigma_1)^{-1} \delta(E_0 - E).$$

Combining these two expressions, we get the following track length formula for the shower as a whole

$$z_\pi(E_0, E) = \alpha E_0/E^2 + (\sigma_1)^{-1} \delta(E_0 - E). \dots\dots\dots (3.11)$$

Friedman (1949) shows that the coefficient of E_0/E^2 in (3.11) must equal the coefficient of the delta function, that is, if we insist on using the very rough approximate form (3.9) at all, then we must make the choice

$$\sigma_1 = \alpha^{-1}. \dots\dots\dots (3.12)$$

It should be emphasized that (3.11) with the choice (3.12) for σ_1 is the best that can be done for the actual shower if we insist on using (3.9) rather than (3.8) for the end group, but it is by no means a close approximation. The best that can be said for it is that the area under the delta function in (3.11) corresponds approximately to the true area underneath that part of the actual z_π which cannot be represented correctly by (3.10).

The reason for stressing the behaviour of v for E close to E_0 is related to the integral equation (2.18). Let us see what region of E' contributes most to the integral in (2.18). There is a factor $(E_s/2E')^2$ as well as $z(E_0, E')$ which itself is proportional to $(E')^{-2}$. These two factors favour low values of E' . The factor $v(E', E, \theta)$ favours high values of E' , but presumably it does so no more strongly than $z_\pi(E_0, E')$ favours low values of E' . The net result is that the maximum in the integrand of (2.18) occurs for low values of E' , that is, for E' very close to E . Thus approximation (3.11) can be used for $z_\pi(E_0, E')$ and approximation (3.7) (with (3.12) for σ_1) or a slightly improved version of it for $v(E', E, \theta)$.

IV. THE FURRY SHOWER

We can find an exact solution to an approximate model of a cascade first introduced by Furry (1937). We can test the qualitative arguments by applying them to the Furry cascade and comparing with the exact solution. It should be noted that there is no qualitative difference between the Furry and actual cascades as far as the angular development of the shower is concerned.

The Furry model contains only one kind of particle which can split into two particles of the same kind. The probability $\sigma_f(E, E')$ for a splitting event in which one of the emerging particles has energy E' , the other energy $E - E'$, is

$$\sigma_f(E, E') = 1/E, \dots\dots\dots (4.1)$$

and the diffusion equation for the average number of particles is

$$\frac{\partial \varphi_{\pi}}{\partial t} = 2 \int_E^{E_0} \varphi_{\pi}(E_0, E', t, \theta) \left(\frac{1}{E'} \right) dE' - \varphi_{\pi} + \left(\frac{E_s}{2E} \right)^2 \frac{\partial^2 \varphi_{\pi}}{\partial \theta^2} + \delta(E_0 - E) \delta(t) \delta(\theta). \quad (4.2)$$

This equation replaces (2.4) and (2.5) for the actual shower. By integrating over t from 0 to infinity we get the track length equation which replaces (2.16). We are most interested in the analogue of (2.18). Since (2.18) does not contain any reference to the photons in the true cascade, it actually holds also for the Furry shower, without modification. The only change is the expression for z_{π} . The longitudinal track length in the Furry shower was found by Nordsieck, Lamb, and Uhlenbeck (1940). They give

$$z_{\pi}(E_0, E) = 2E_0/E^2 + \delta(E_0 - E). \quad (4.3)$$

According to the qualitative arguments of Section III, the end group (corresponding to the $\delta(E_0 - E)$ in (4.3)) should have a track length given by $v^{(0)}$, (3.7), with $\sigma_1 = 1$; low energy particles should have a track length given by $v = (E/E_s) z_{\pi} f(u)$, according to (3.2) and (3.3). We therefore make the initial assumption

$$v(E_0, E, \theta) = v^{(0)}(E_0, E, \theta) + (E/E_s)(2E_0/E^2) f(E\theta/E_s), \quad (4.4)$$

where f is so far an unknown function. We substitute (4.3) and (4.4) into the integro-differential equation (2.18) to get an equation for f . We make use of the fact that

$$\frac{d^2 \exp(-a|\theta|)}{d\theta^2} = a^2 \exp(-a|\theta|) - 2a\delta(\theta) \quad (4.5)$$

in order to get the following differential equation for $f(u) = f(E\theta/E_s)$:

$$\frac{d^2 f}{du^2} - 4f(u) = -4 \exp(-2|u|). \quad (4.6)$$

The fact that this is an equation not involving the energy variables E , E_0 explicitly shows that the initial assumption (4.4) is consistent with the equations defining the Furry cascade. The differential equation (4.6) is solved easily, subject to the boundary conditions that $f(u)$ vanish for $u = +\infty$ and $u = -\infty$. The solution is

$$f(u) = \frac{1}{2} e^{-2|u|} + |u| e^{-2|u|}. \quad (4.7)$$

Expression (4.4) with (4.7) for $f(u)$ is an exact solution for the Furry cascade in the Landau approximation. Since the Ansatz (4.4) was constructed on the basis of the qualitative arguments in Section III, these qualitative arguments are thereby strengthened considerably, and we can have some confidence in their application to the actual shower.

It is worth remarking that, unlike the expression (3.7) for the angular distribution of the end group of particles, the angular distribution for energies $E < E_0$ (4.7) does not have a cusp at $u = 0$. Correspondingly, the distribution in $U = E\Theta/E_s$ ($\Theta = (\theta_x^2 + \theta_y^2)^{1/2}$) implied by (4.7) is finite for all U , indeed it is

$$F(U) = (2U/\pi) K_1(2U). \quad (4.8)$$

Thus, in the Landau approximation at least, the analytic behaviour of the end group near $\theta=0$ does not give a reliable guide to the analytic behaviour of the desired solution.

V. THE ACTUAL SHOWER: ROUGH SOLUTION

In order to get a first orientation about the angular distribution in the actual shower, we use the integro-differential equation (2.18) with the approximation (3.11) for $z_\pi(E_0, E)$ and with the following Ansatz for $v(E_0, E, \theta)$:

$$v(E_0, E, \theta) = v^{(0)}(E_0, E, \theta) + (E/E_s)(\alpha E_0/E^2)f(\theta E/E_s). \quad \dots\dots (5.1)$$

In this expression $v^{(0)}$ is taken from (3.7) with the choice $\sigma_1=1/\alpha$ for the "effective radiative cross section" (see the discussion in connexion with formula (3.12)). $f(u)$ is an unknown function.

Substitution into (2.18) and performance of the integration over E' leads to the following result:

$$f(u) = (\alpha/8)[1 + (E^2/E_0^2)]f''(u) + \exp(-2\alpha^{-1}|u|). \quad \dots (5.2)$$

The presence of the ratio $(E/E_0)^2$ in this equation means that the Ansatz (5.1) is not consistent with the shower equations. When $E \ll E_0$, however, the coefficient of f'' becomes approximately independent of energy. We propose to take this limit in solving the equation. In that case the solution of (5.2) is, with

$$\mu = \alpha^{-1} = 1.46, \quad \dots\dots\dots (5.3)$$

given by

$$f(u) = 2^{1/2}\mu[2^{1/2}\exp(-2\mu|u|) - \exp(-8^{1/2}\mu|u|)]. \quad \dots (5.4)$$

This expression has the property that $f(0)$ is finite and $f'(0)=0$, just as the Furry shower $f(u)$, (4.7). The integral of $f(u)$ over all values of u is 1, as it should be. This relation depends upon the choice (3.12) for the effective cross section σ_1 . Any other choice would have led to an incorrect value for $\int_{-\infty}^{\infty} f(u)du$.

Since σ_1 was chosen on the basis of considerations about the longitudinal development of the shower, the fact that this same σ_1 makes sense for the angular development is encouraging.

(5.4) can not be an exact solution. We have used an approximate formula for z_π , expression (3.10), and we have ignored a term in the differential equation (5.2). Direct comparison of the moments of (5.4) with the precise moments (in the Landau approximation) shows the same thing; the moments disagree, increasingly so as we go to higher and higher moments. (These moments are given in Section VI, Table 3.)

VI. THE ACTUAL SHOWER: ITERATED SOLUTION

The best way to test the accuracy of an approximate solution of some complicated equation is to compare it with the exact solution. If no exact solution is available, the next best thing is to iterate on the approximate solution in some systematic manner. If the iterated function does not differ greatly from the first approximate function, then the difference between the two is

probably of the order of the absolute error of the original. In addition, the iterated function is presumably a better approximation to the truth than the original.

The best method to judge the accuracy of the iterated function is to perform a second iteration, with the first iterate as trial function. As will be seen later, this is practically impossible here. However, this does not mean that we have no way of testing the accuracy of the iterated function. We shall perform this test by comparing the moments of the iterated function with the exact (in Landau approximation) moments.

Friedman (1949) has formulated a variation-iteration technique for the solution of equations such as (2.16). The Friedman method is discussed in Appendix III. This method has two major advantages: (1) the normalization of the iterated function is independent of the normalization of the trial function, (2) the error in the iterated function is of the order of magnitude of the *square* of the error in the trial function (this latter property is implied by the term "variation method").

Let $v_0(E_0, E, \theta)$ be a trial solution of the integro-differential equation (2.16). The iterated solution in the Friedman scheme is then given by

$$v_1(E_0, E, \theta) = \frac{[v_0(E_0, E, \theta)]^2}{\int_{E_0}^E dE' \int_{-\infty}^{\infty} d\theta' [Fv_0(E_0, E', \theta') - (E_s/2E')^2 \partial^2 v_0 / \partial \theta'^2] v_0(E', E, \theta - \theta')} \quad (6.1)$$

It is easily seen from (2.16) that the choice $v_0 = v$ (trial function = exact solution) leads to $v_1 = v$ also. It is shown in Appendix III that the choice $v_0 = v + E$ leads to $v_1 = v + \text{Order}(E^2)$, E being the error of the trial function.

We have carried out the iteration (6.1) explicitly, using as our trial function the result of Section V; since the normalization of v_0 in (6.1) is of no importance, we multiply (5.1) by the constant $\mu E_s/2$; we also use the notation

$$x = 2\mu u = 2\mu E\theta/E_s \quad (6.2)$$

We shall use x as our angle variable consistently. With this notation, the trial function becomes

$$v_0(E_0, E, x) = (E_0/E) \left\{ \exp(-|x|) - 2^{-1/2} \exp(-2^{1/2}|x|) \right\} + \frac{1}{2} E_0 \delta(E_0 - E) \exp(-E_0|x|/E), \quad \left. \vphantom{\exp(-|x|)} \right\} \\ v_0 = 0, \quad \text{when } E > E_0. \quad (6.3)$$

The operator F , (2.17), can be found explicitly from (1.1) and (2.1)–(2.3). We shall work with a finite but small value of ε in (2.1), and let ε approach 0 at the very end of the calculation. By interchanging orders of integration in the operator product BC , we get the following explicit form for the operator F acting on an arbitrary function $f(E)$

$$Ff = \ln(E/\varepsilon)f(E) - \int_{E+\varepsilon}^{\infty} f(E')(E' - E)^{-1} dE' - 2 \int_E^{\infty} f(E')(E^{-1} - E'^{-1}) dE'. \quad (6.4)$$

We now apply this operator F to the function $f=(E_0/E)\exp(-|x|)$. Since x is defined by (6.2) so as to contain the energy variable E , we get

$$f(E')=(E_0/E')\exp(-E'|x|/E), \quad f=0 \text{ for } E'>E_0, \quad \dots \quad (6.5)$$

and

$$Ff=(E_0/E)\left\{\ln(E/\varepsilon)\exp(-|x|)-\int_{1+(E_0/E)}^{E_0/E}\frac{\exp(-y|x|)}{y(y-1)}dy\right. \\ \left.-2\int_1^{E_0/E}y^{-1}(1-y^{-1})\exp(-y|x|)dy\right\}=(E_0/E)G(x,E_0/E). \quad \dots \quad (6.6)$$

The function G defined by (6.6) can be expressed in closed form in terms of the exponential integral. We use the standard notation (Jahnke and Emde 1943)

$$-\text{Ei}(-x)=\int_x^\infty\frac{e^{-y}}{y}dy, \quad x>0, \quad \dots \dots \dots (6.7)$$

and the Euler constant

$$\gamma=1.7811, \quad \ln(\gamma)=C=0.57722. \quad \dots \dots \dots (6.8)$$

We then get, after some manipulation,

$$G(x,y)=[\ln(\gamma|x|)+2-\text{Ei}(|x|-|xy|)]e^{-|x|} \\ +(1+2|x|)[\text{Ei}(-|x|)-\text{Ei}(-|xy|)]-2y^{-1}e^{-|xy|}. \quad \dots \quad (6.9)$$

We introduce the notation g for the following function :

$$Fv_0=(E_0/E)g(x,E_0/E), \quad \dots \dots \dots (6.10)$$

to get

$$g(x,y)=G(x,y)-2^{-\frac{1}{2}}G(2^{\frac{1}{2}}x,y)+\frac{1}{2}\delta(y-1)\ln(E_0/\varepsilon)e^{-|xy|} \\ +[y^{-1}-1-\frac{1}{2}(y-1)^{-1}]e^{-|xy|}. \quad \dots \dots \dots (6.11)$$

The term with $\ln(E_0/\varepsilon)$ comes from the bremsstrahlung cross section divergence. Corresponding to the cut-off we have used for the bremsstrahlung cross section, we must assume that there are no electrons in the shower of energy between E_0 and $E_0-\varepsilon$. Thus the formulae written so far are not meant to apply within that region. In particular, this is important for the last term of (6.11), which becomes infinite at $y=1$, and would give a logarithmic singularity if integrated straightforwardly, in a later stage of this calculation (formulae (6.16) and (6.18)).

We also need the second term on the left-hand side of (2.16). We introduce the function h by

$$-(E_s/2E)^2\partial^2v_0/\partial\theta^2=-\mu^2\partial^2v_0/\partial x^2=+(\mu^2E_0/E)h(x,E_0/E). \quad \dots \quad (6.12)$$

By using relation (4.5) we find

$$h(x,y)=\delta(y-1)\delta(x)-\delta(y-1)\exp(-|xy|)-\exp(-|x|)+2^{\frac{1}{2}}\exp(-2^{\frac{1}{2}}|x|). \\ \dots \dots \dots (6.13)$$

We now substitute these expressions into the denominator of (6.1). We introduce, instead of θ' , the variable of integration

$$x'=2\mu E\theta'/E_s, \quad \dots \dots \dots (6.14)$$

and we let D stand for the denominator of (6.1). We use the variable of integration $y=E'/E$ to get the result

$$D=(E_0 E_0/2\mu E)\int_1^{E_0/E} dy \int_{-\infty}^{+\infty} dx' [g(xy, E_0/E) + \mu^2 h(xy, E_0/E)] \\ \times [\exp(-|x-x'|) - 2^{-\frac{1}{2}} \exp(-2^{\frac{1}{2}}|x-x'|) + (1/2y)\delta(y-1) \exp(-|x-x'|)]. \quad (6.15)$$

It is very convenient to do the integration over energy (over y) before the integration over the angle variable x' . We introduce the functions

$$\left. \begin{aligned} g_1(x, y) &= \int_1^y dy' g(xy', y/y'), \\ h_1(x, y) &= \int_1^y dy' h(xy', y/y'), \end{aligned} \right\} \quad (6.16)$$

to get the following expression for the denominator D :

$$D=(E_0 E_0/2\mu E)\int_{-\infty}^{+\infty} dx' [g(x', E_0/E) + \mu^2 h(x', E_0/E)]^{\frac{1}{2}} \exp(-|x-x'|) \\ + (E_0 E_0/2\mu E)\int_{-\infty}^{+\infty} dx' [g_1(x', E_0/E) + \mu^2 h_1(x', E_0/E)] [\exp(-|x-x'|) \\ - 2^{-\frac{1}{2}} \exp(-2^{\frac{1}{2}}|x-x'|)]. \quad (6.17)$$

Explicit evaluation of the functions g_1 and h_1 defined by (6.16) is possible, and gives the following results:

$$g_1(x, y) = G_1(x, y) - 2^{-\frac{1}{2}} G_1(2^{\frac{1}{2}}x, y) + \frac{1}{2} e^{-|xy|} [y \ln(y/(y-1)) + 1 - y^{-1}], \quad (6.18)$$

$$G_1(x, y) = [x^{-1} \ln(\gamma x) - 1] e^{-x} - (x+1+x^{-1}) \text{Ei}(-x) \\ - [x^{-1} \ln(\gamma xy) - y] e^{-xy} + (x+1+x^{-1}) \text{Ei}(-xy) \\ + x^{-1} \ln[\gamma x(y-1)] e^{-xy} - x^{-1} e^{-x} \text{Ei}(x-xy) \\ - (y-y^{-1}) e^{-xy}, \quad \text{for } x > 0, \quad (6.19)$$

$$G_1(-x, y) = G_1(x, y),$$

$$h_1(x, y) = |x|^{-1} [\exp(-|xy|) - \exp(-|x|) - \exp(-2^{\frac{1}{2}}|xy|) + \exp(-2^{\frac{1}{2}}|x|)] \\ - \frac{1}{2} y \exp(-|xy|) + \delta(x). \quad (6.20)$$

We are now at the stage where we can make good use of the fact that we want the final expressions only in the limit $y=E_0/E \rightarrow \infty$. We therefore define limiting functions

$$\bar{g}(x) = \lim_{y \rightarrow \infty} g(x, y), \quad (6.21)$$

and similar expressions for all the other functions involved. Considerable care must be exercised in taking this limit. An expression such as $y \exp(-|xy|)$

becomes zero in this limit for all non-zero values of x , but can not therefore be ignored. The fact that

$$\int_{-\infty}^{+\infty} y \exp(-|xy|) dx = 2$$

is independent of y indicates that such a term gives rise to a delta function contribution in the limiting process (6.21), in this particular case equal to $2\delta(x)$. The delta function contributions are very important. Inspection of (6.17) shows that $\delta(x)$ terms in \bar{g} or \bar{h} give rise to terms proportional to $\exp(-|x|)$; there are no such terms, actually; but a $\delta(x)$ term in \bar{h}_1 , which does occur, gives rise to a term proportional to the original trial structure function (5.4). If (5.4) had been an exact solution, the denominator D would be in its entirety proportional to (5.4). Thus, the more nearly the trial function is equal to the exact solution, the more nearly we have $\bar{g} + \mu^2 \bar{h} = 0$ and $\bar{g}_1 + \mu^2 \bar{h}_1$ equal to a constant times $\delta(x)$.

The terms proportional to $\delta(y-1)$ in (6.11) and (6.13) can be ignored in the limit $y \rightarrow \infty$. There is one remark to be made, however. The $\delta(y-1)$ term in $g(x,y)$ (6.11) is multiplied by the constant $\ln(E_0/\varepsilon)$; if we go to the limit $\varepsilon \rightarrow 0$, corresponding to the actual (supersimplified) bremsstrahlung cross section, this term becomes infinite. It contributes only at the initial energy $E = E_0$ (i.e. $y=1$), and the fact that the denominator of (6.1) has an infinite contribution proportional to $\delta(E_0 - E)$ indicates that the iterated function $v_1(E_0, E, \theta)$, unlike the trial function v_0 , contains no term proportional to $\delta(E_0 - E)$. This is of course to be expected of the true solution v , and the fact that the very first iteration already gives this result is encouraging.

We now write down the barred functions:

$$\bar{g}(x) = \bar{G}(x) - 2^{-\frac{1}{2}} \bar{G}(2^{\frac{1}{2}}x), \quad \dots \quad (6.22)$$

$$\bar{G}(x) = [\ln(\gamma |x|) + 2]e^{-|x|} + (1 + 2|x|)\text{Ei}(-|x|), \quad \dots \quad (6.23)$$

$$\bar{h}(x) = -\exp(-|x|) + 2^{\frac{1}{2}} \exp(-2^{\frac{1}{2}}|x|), \quad \dots \quad (6.24)$$

$$\bar{g}_1(x) = \bar{G}_1(x) - 2^{-\frac{1}{2}} \bar{G}_1(2^{\frac{1}{2}}x), \quad \dots \quad (6.25)$$

$$\bar{G}_1(x) = [|x|^{-1} \ln(\gamma |x|) - 1]e^{-|x|} - (|x| + 1 + |x|^{-1})\text{Ei}(-|x|), \quad \dots \quad (6.26)$$

$$\bar{h}_1(x) = -|x|^{-1}[\exp(-|x|) - \exp(-2^{\frac{1}{2}}|x|)] + \ln(2)\delta(x). \quad \dots \quad (6.27)$$

It will be noticed that, although there are several "delta function terms" in $G_1(x,y)$ (6.19) these actually cancel out.*

In order to exhibit the contribution of the delta function term in $\bar{h}_1(x)$ more explicitly, we write

$$\left. \begin{aligned} \bar{h}_1(x) &= \bar{h}_2(x) + \ln(2)\delta(x), \\ \bar{h}_2(x) &= -|x|^{-1}[\exp(-|x|) - \exp(-2^{\frac{1}{2}}|x|)]. \end{aligned} \right\} \dots \quad (6.27a)$$

* The delta function terms were found by a more indirect method by Kalos (1952) and there appeared to be a very small delta function contribution in $\bar{G}_1(x)$. This was the result of round off errors in the numerical calculations. No significant change is introduced thereby into the final results; if anything, the iterated function is improved a little by this correction.

The denominator of (6.1) in the limit $E \ll E_0$ is then given by (6.17) with the barred functions \bar{g} etc. instead of the original functions g etc. We now transform the integrals slightly, as follows :

$$\int_{-\infty}^{+\infty} dx' \bar{g}(x') \exp(-|x-x'|) = 2 \int_0^{\infty} dx' \bar{g}(x') K(x, x'), \quad \dots (6.28)$$

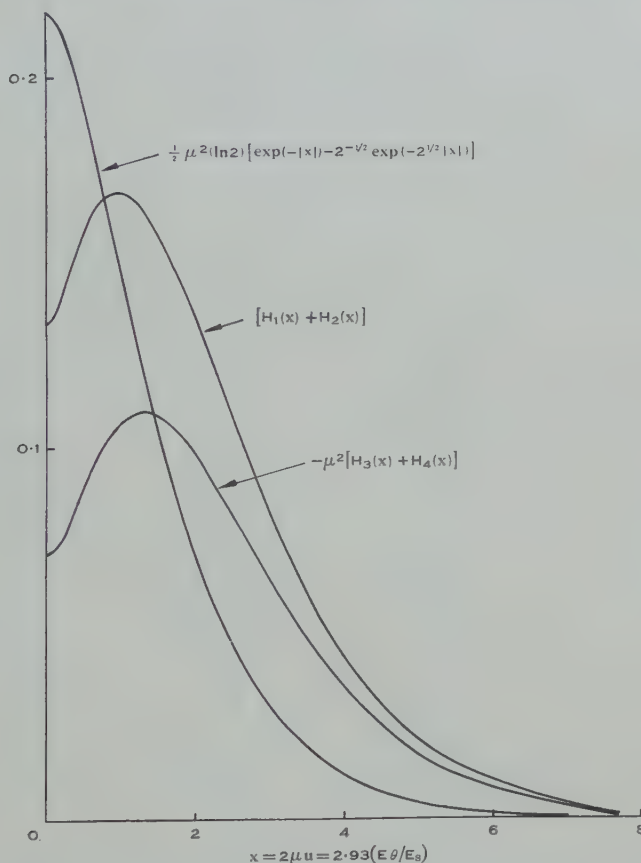


Fig. 1.—Intermediate quantities in the iteration of the track length angular distribution. The three quantities shown here add up to form the denominator of equation (6.35); notice the reversed sign on $-\mu^2(H_3+H_4)$. The contribution of the “delta-function term”, i.e. the term proportional to the trial function, is very appreciable.

where the kernel $K(x, x')$ is given by

$$\left. \begin{aligned} K(x, x') &= \exp(-x) \cosh x', & \text{if } x > x', \\ &= \exp(-x') \cosh x, & \text{if } x < x'. \end{aligned} \right\} \dots (6.29)$$

We then collect together the terms with the same kernel, to get

$$\begin{aligned} D &= (E_s E_0 / \mu E) [H_1(x) + H_2(x) + \mu^2 H_3(x) + \mu^2 H_4(x) \\ &\quad + \tfrac{1}{2} \mu^2 (\ln 2) \{ \exp(-|x|) - 2^{-1/2} \exp(-2^{1/2}|x|) \}], \quad \dots (6.30) \end{aligned}$$

where

$$H_1(x) = \int_0^\infty dx' [\tfrac{1}{2}\bar{g}(x') + \bar{g}_1(x')] K(x, x'), \quad \dots\dots (6.31)$$

$$H_2(x) = -2^{-\frac{1}{2}} \int_0^\infty dx' \bar{g}_1(x') K(2^{\frac{1}{2}}x, 2^{\frac{1}{2}}x'), \quad \dots\dots (6.32)$$

$$H_3(x) = \int_0^\infty dx' [\tfrac{1}{2}\bar{h}(x') + \bar{h}_2(x')] K(x, x'), \quad \dots\dots (6.33)$$

$$H_4(x) = -2^{-\frac{1}{2}} \int_0^\infty dx' \bar{h}_2(x') K(2^{\frac{1}{2}}x, 2^{\frac{1}{2}}x'). \quad \dots\dots (6.34)$$

TABLE 1
INTERMEDIATE QUANTITIES IN THE ANGULAR DISTRIBUTION ITERATION

x	$H_1(x)$	$H_2(x)$	$H_3(x) + H_4(x)$
0	0.53779	-0.40389	-0.03337
0.1	0.53182	-0.39599	-0.03393
0.2	0.52163	-0.38017	-0.03534
0.3	0.50818	-0.36057	-0.03727
0.4	0.49256	-0.33909	-0.03944
0.5	0.47548	-0.31688	-0.04166
0.6	0.45744	-0.29466	-0.04379
0.7	0.43881	-0.27290	-0.04573
0.8	0.41987	-0.25190	-0.04742
0.9	0.40085	-0.23187	-0.04880
1.0	0.38192	-0.21292	-0.04987
1.2	0.34486	-0.17847	-0.05105
1.4	0.30947	-0.14863	-0.05102
1.6	0.27623	-0.12316	-0.04996
1.8	0.24540	-0.10165	-0.04807
2.0	0.21710	-0.08363	-0.04557
3.0	0.11109	-0.03028	-0.02965
4.0	0.05394	-0.01059	-0.01623
5.0	0.02499	-0.00368	-0.00804
7.0	0.00485	-0.00042	-0.00166
9.0	0.00087	-0.00005	-0.00030

Combining equations (6.1), (6.3), (6.30), and (3.2), we get the following expression for the angular structure function of the shower at energies $E \ll E_0$ (with $x = 2\mu u = 2\mu E\theta/E_s$)

$$f(E\theta/E_s) = \frac{\mu^3 \{ \exp(-|x|) - 2^{-\frac{1}{2}} \exp(-2^{\frac{1}{2}}|x|) \}^2}{\bar{H}_1(x) + H_2(x) + \mu^2 [H_3(x) + H_4(x)] + \frac{1}{2}\mu^2 \ln 2 \{ \exp(-|x|) - 2^{-\frac{1}{2}} \exp(-2^{\frac{1}{2}}|x|) \}} \quad \dots\dots\dots (6.35)$$

Before going into the evaluation of the integrals (6.31) to (6.34), we point out one general property of them, namely, the behaviour of these functions

near $x=0$. By differentiating $H_1(x)$, say, under the integration sign in (6.31), we can show immediately that the derivative of $H_1(x)$ vanishes at $x=0$. The same holds for the other integrals. The $\delta(x)$ contribution gives rise to a constant times the trial function (5.4), and this trial function itself has zero derivative at $x=0$. We therefore conclude that the whole denominator D , and hence also the iterated function v_1 of (6.1), has zero derivative at $x=0$. There is no singularity

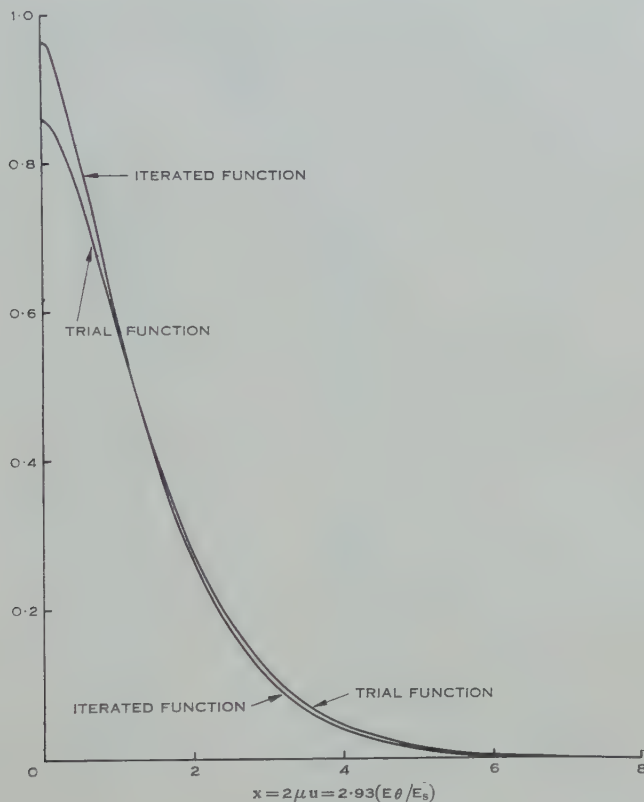


Fig. 2.—The trial function and the iterated function are rather close to each other over most of the range, indicating that the iterated function is likely to be very close to the true mathematical solution of the equations.

of the angular distribution function of monoenergetic electrons in the Landau approximation. This result is in complete agreement with the results of Belenky (1944) and of Molière (1946). The integrals H_3 and H_4 can be done analytically. The results are, for $x > 0$:

$$H_3(x) = \frac{1}{2}e^{-x}\left\{\frac{1}{2}(3-x) + \text{Ei}[(1-2^{\frac{1}{2}})x] - \ln(\frac{1}{2}\gamma x)\right\} \\ + \frac{1}{2}e^{+x}\{\text{Ei}(-2x) - \text{Ei}[-(1+2^{\frac{1}{2}})x]\}, \quad \dots\dots\dots (6.36)$$

$$H_4(x) = -8^{-\frac{1}{2}}\exp(-2^{\frac{1}{2}}x)\{\text{Ei}[(2^{\frac{1}{2}}-1)x] + \ln(8^{-\frac{1}{2}}\gamma x)\} \\ + 8^{-\frac{1}{2}}\exp(+2^{\frac{1}{2}}x)\{\text{Ei}(-8^{\frac{1}{2}}x) + \text{Ei}[-(1+2^{\frac{1}{2}})x]\}. \quad \dots\dots\dots (6.37)$$

The integrals (6.31) and (6.32) are best done numerically, and have been done in this fashion. Values of $H_1(x)$, $H_2(x)$, and of the sum H_3+H_4 are given in Table 1. These functions are plotted in Figure 1. There we have plotted also the contribution from the delta function term in \bar{h}_1 so as to show the tremendous importance of this delta function contribution. It will also be noticed that the sum (H_1+H_2) nearly cancels $\mu^2(H_3+H_4)$ for small values of x .

TABLE 2
ANGULAR STRUCTURE FUNCTION: THE TRIAL FUNCTION AND THE
ITERATED FUNCTION; $x=2\mu E\theta/E_s=2\mu u$

x	$f_{\text{trial}}(x)$	$f_{\text{iter}}(x)$
0	0.85794	0.96235
0.1	0.85234	0.95220
0.2	0.83726	0.92283
0.3	0.81489	0.88584
0.4	0.78709	0.84446
0.5	0.75538	0.80065
0.6	0.72100	0.75584
0.7	0.68492	0.71082
0.8	0.64800	0.66646
0.9	0.61087	0.62295
1.0	0.57404	0.58089
1.2	0.50277	0.50116
1.4	0.43634	0.42993
1.6	0.37585	0.36603
1.8	0.32175	0.30997
2.0	0.27400	0.26117
3.0	0.11607	0.10437
4.0	0.046414	0.03882
5.0	0.017978	0.01373
7.0	0.0025671	0.00159
9.0	0.00035534	0.000171

The trial function and the iterated function are compared in Table 2 and in Figure 2. The change produced by the iteration is not very large except in the "tail" of the distribution function (large values of x), where the iterated function is smaller than the trial function, as it should be.

We now examine the moments of the trial function, iterated function, and the "exact" moments. These moments are given in Table 3. Two sets of exact moments are given there, one for supersimplified cross sections, the other for conventional cross sections. We see immediately that the iterated function is a great improvement over the trial function. For example, the second moment of the trial function is in error by 17 per cent., the second moment of the iterated function by less than 3 per cent. The improvement is even bigger for the higher moments.

Next, let us compare the iterated function with the true function in the approximation we have been using (supersimplified cross sections) through their

moments. The second moment is too large by 3 per cent., the fourth moment by 7 per cent., the sixth moment by 14 per cent. The main contribution to the fourth moment comes from values of x near 4, that is, values of u near $4/3$, and to the sixth moment from x near 6 or u near 2. We conclude that *the iterated*

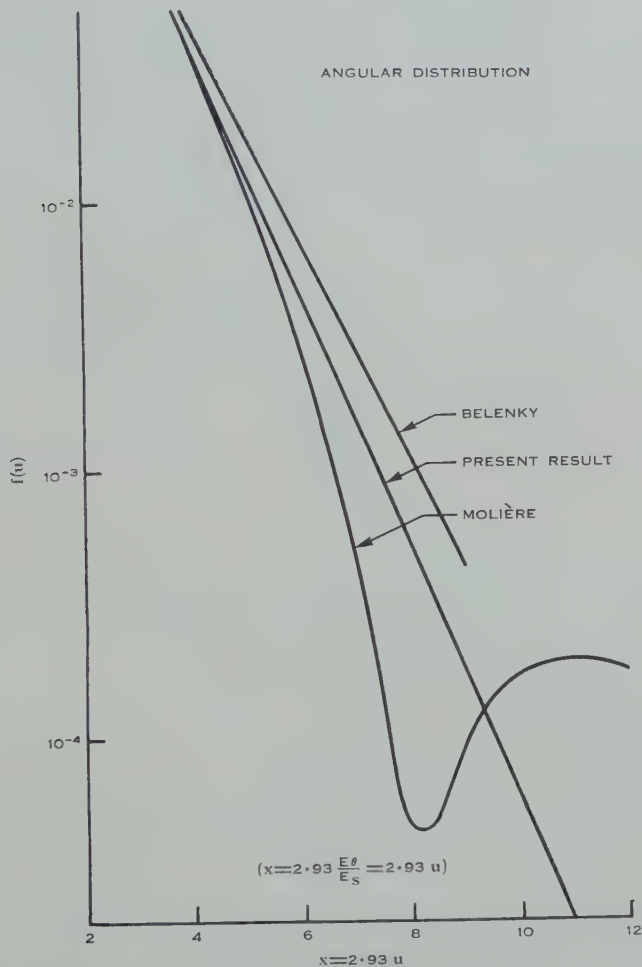


Fig. 3.—Comparison with other work. The results of Molière and Belenky differ from our iterated function out in the tail. A comparison with the exact moments indicates that our iterated function is the best of the three. No special significance attaches to this region of angles, since the basic approximations (especially the Landau approximation of pure multiple scattering) break down there.

function is within 10 per cent. of the mathematically correct solution of our equations for values of $u = E\theta/E_s < 1.5$.

Our equations contain several approximations within them. One of these is the use of supersimplified cross sections. When we compare the moments of

our iterated function with the exact moments using conventional cross sections, the agreement is of course not nearly as good. The second moment is too large by 8 per cent., the fourth moment by 21 per cent., and the sixth moment by 41 per cent. While it would be in principle possible to use the conventional rather than the supersimplified cross sections for the operator F in the denominator of the iteration equation (6.1), we feel that the additional work involved is hardly warranted by the present experimental data on the angular

TABLE 3
MOMENTS OF THE ANGULAR DISTRIBUTION, IN LANDAU APPROXIMATION

n	$f_n = \langle (E\theta/E_s)^n \rangle_{\text{ave}}$			
	Trial	Iterated	Exact	
			Supersimplified Cross Sections	Conventional Cross Sections
0	1.00	1.01	1.00	1.00
2	0.350	0.309	0.300	0.285
4	0.571	0.433	0.406	0.360
6	2.138	1.362	1.191	0.970
8	14.416	7.654	5.940	4.89

distribution of monoenergetic electrons. The moments make it likely that the iterated function v_1 is within 10 per cent. of the mathematically correct solution with conventional cross sections for values of $u < 0.8$.

Next, our function is in error by an unknown amount because of the use of the Landau approximation. We intend to perform the iteration once more, using the correct integral operator for Coulomb scattering rather than the operator $(E_s/2E)^2 \partial^2 / \partial \theta^2$ in the denominator of (6.1). The discussion of the errors produced by the Landau approximation is reserved for a later publication.

Last, we compare our iterated function with the results of Belenky (1944) and Molière (1946). This is done in Figures 3 and 4. We see that the three functions agree very well indeed for small values of x . The methods of calculation used by these authors were different from each other and different from ours. Molière used supersimplified cross sections, and Belenky used the cross sections of the Tamm-Belenky theory, which are closely related to the supersimplified cross sections. Both authors also used the Landau approximation for the Coulomb scattering. Thus it is not surprising that their results should agree so well with ours. A comparison of the moments of the various functions with the exact moment shows that for larger values of x , where the three functions disagree more and more, our v_1 is the closest approximation to the mathematically exact solution. However, no great stress should be laid on this point, since the Landau approximation has in any case no claim to validity for $u > 1$ (for $x > 3$). Molière's angular distribution function has a minimum at $x=8$ ($u=2.7$), and the corresponding distribution-in- U ($U=E\Theta/E_s$) actually becomes negative there, but it is apparent from Figure 3 that this is way out in the "tail" of the

function, and hence the importance of this error in Molière's calculation has been overestimated in the past.*

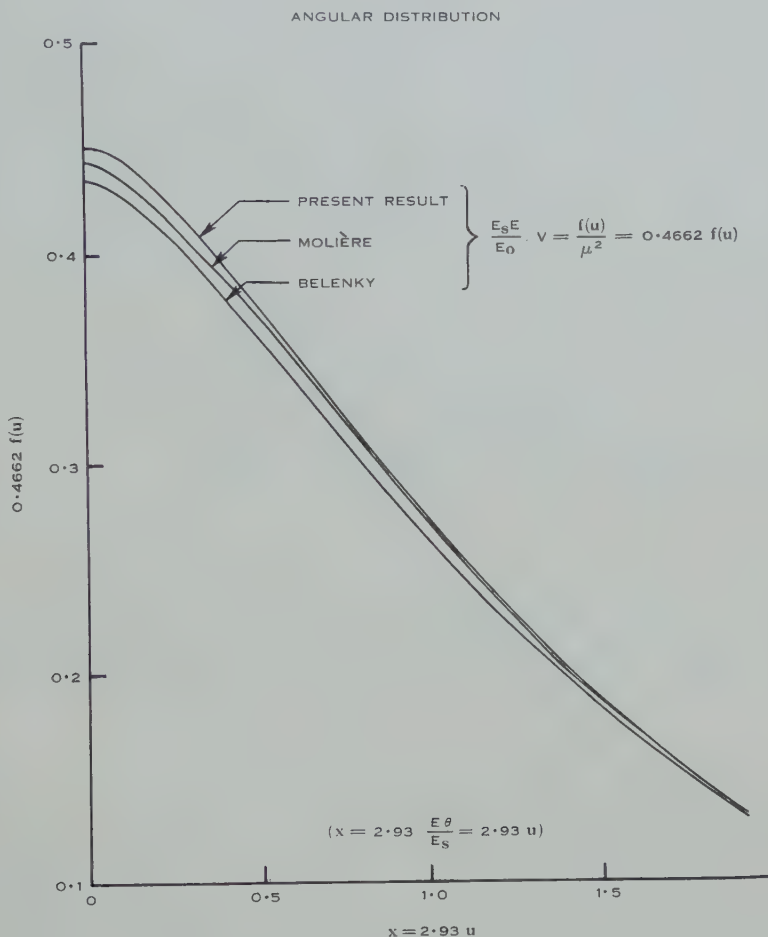


Fig. 4.—Comparison with other work. The results of Molière and Belenky agree closely with our iterated function in the significant region of angles. (Notice that the ordinate scale does not start at 0.) The dominant contribution to the second moment of f comes from the neighbourhood of $x=2$, at the extreme right of the figure. We conclude that it is impractical to deduce the behaviour of $f(u)$ for small values of u from a study of the moments of f .

* Blatt (1949). However, the doubts concerning the accuracy of Molière's *lateral* structure function are not decreased appreciably by the agreement of the angular structure function. The lateral structure function is very much harder to determine. The only improvements on Molière's work published to date have been concerned with the moments of the lateral distribution. These moments disagree with Molière's moments quite badly even in the Landau approximation, for example the sixth moment of Molière is too small by a factor 4.65, the 10th moment of Molière is too small by a factor 1000. These factors get worse if the Landau approximation is dropped.

VII. REFERENCES

- BELENKY, S. (1944).—*J. Phys. U.S.S.R.* **8** : 347.
 BHABHA, H. J., and HEITLER, W. (1937).—*Proc. Roy. Soc. A* **159** : 432.
 BLATT, J. M. (1949).—*Phys. Rev.* **75** : 1584.
 COCCONI, G. (1954).—*Phys. Rev.* **93** : 646.
 EYGES, L., and FERNBACH, S. (1951).—*Phys. Rev.* **82** : 23.
 FRIEDMAN, F. L. (1949).—MIT Lab. for Nuclear Science and Engineering Tech. Rep. No. 31.
 FURRY, W. H. (1937).—*Phys. Rev.* **52** : 569.
 GREEN, H. S., and MESSEL, H. (1952).—*Phys. Rev.* **88** : 331.
 JAHNKE, E., and EMDE, F. (1943).—"Tables of Functions." (Dover Publications : New York.)
 KALOS, M. H. (1952).—On the lateral structure of electronic showers in air. Thesis, University of Illinois.
 LANDAU, L. (1940).—*J. Phys. U.S.S.R.* **3** : 237.
 MOLIÈRE, G. (1946).—"Cosmic Radiation." (Ed. W. Heisenberg.) Ch. 3. (Dover Publications : New York.)
 NORDHEIM, L. W. (1952).—*Z. Phys.* **133** : 94.
 NORDSIECK, A. T., LAMB, W. E., and UHLENBECK, G. E. (1940).—*Physica* **7** : 344.
 ROBERG, J., and NORDHEIM, L. W. (1949).—*Phys. Rev.* **75** : 444.
 ROSSI, B. (1952).—"High Energy Particles." (Prentice Hall Inc. : New York.)
 ROSSI, B., and GREISEN, K. (1941).—*Rev. Mod. Phys.* **13** : 240.
 SNYDER, H. S., and SCOTT, W. T. (1949).—*Phys. Rev.* **76** : 220.
 WATSON, G. N. (1948).—"Bessel Functions." (Cambridge Univ. Press.)

APPENDIX I

Moments of the Angular Distribution

We start from the Nordheim recursion formula (2.12). We take the Mellin and Laplace transform on both sides, and we assume that the dependence of the transform $\pi_n(E_0, s, \lambda)$ on the initial energy E_0 is given by

$$\int_0^{E_0} dE E^s \int_0^\infty dt e^{\lambda t} \pi_n(E_0, E, t) = (E_0)^{s-n} M_n(s, \lambda). \quad \dots \quad (\text{A1})$$

We then get from (2.12), by inverting orders of integration,

$$M_n(s, \lambda) = n(n-1)(\tfrac{1}{2}E_s)^2 M_0(s-n, \lambda) M_{n-2}(s, \lambda). \quad \dots \quad (\text{A2})$$

The fact that E_0 does not appear in (A2) shows that the assumption (A1) was correct. The solution of this recurrence relation is immediate, giving

$$M_n(s, \lambda) = n! (\tfrac{1}{2}E_s)^n M_0(s, \lambda) M_0(s-2, \lambda) M_0(s-4, \lambda) \dots M_0(s-n, \lambda). \quad \dots \quad (\text{A3})$$

The Laplace Mellin transform $M_0(s, \lambda)$ refers to the zeroth moment of the angular problem, that is, to the longitudinal distribution function $\pi_0 = \pi$ of Rossi and Greisen (1941). By comparison with equation (2.33) of Rossi and Greisen we see that

$$M_0(s, \lambda) = \frac{\sigma_0 + \lambda}{[A(s) + \lambda](\sigma_0 + \lambda) - B(s)C(s)}. \quad \dots \quad (\text{A4})$$

(A3) and (A4) constitute the complete solution to the moment problem in the Landau approximation. It will be noticed that the depth transform variable λ enters only in a trivial way here, that is, the calculation for arbitrary λ is in no way more difficult than the track length calculation ($\lambda=0$). The inversion of the Laplace transform, to get explicit functions of the depth t , is straightforward although somewhat tedious. On the other hand, the inversion of the Mellin

transform for arbitrary t can be done only by the saddle point method. This is the reason why most published results do not give the depth dependent moments, even though they are in principle just as well known as the track length moments. The depth dependence of $\langle \theta^2 \rangle_{\text{ave}}$ was given in a report by Osborne, Nordheim, and Blatt (unpublished, Echo Lake Conference 1949).

In this paper, we are concerned only with the track length values, so we set $\lambda=0$ consistently. We can then invert the Mellin transform by using the theory of residues. For $\lambda=0$, the pole of $M_0(s,0)$ with the largest real part of s occurs at the point $s=1$. Thus the dominant pole of (A3) occurs at $s=n+1$. The other poles give contributions involving lower powers of (E_0/E) and can therefore be ignored in the limit $E \ll E_0$. We define the n th moment of the angular structure function $f(u)$, (3.2) and (3.3), by

$$f_n = \left\langle \left(\frac{E\theta}{E_s} \right)^n \right\rangle_{\text{ave}} = \left(\frac{E}{E_s} \right)^n \frac{z_{\pi n}(E_0, E)}{z_{\pi}(E_0, E)}, \quad \dots \quad (\text{A5})$$

where $z_{\pi n}$ is the track length n th angular moment, that is,

$$z_{\pi n}(E_0, E) = \int_0^\infty \pi_n(E_0, E, t) dt = \int_{-\infty}^{+\infty} \theta^n v(E_0, E, \theta) d\theta. \quad \dots \quad (\text{A6})$$

This procedure leads to the result, for $E \ll E_0$,

$$f_n = n! 2^{-n} M_0(3,0) M_0(5,0) M_0(7,0) \dots M_0(n+1,0). \quad \dots \quad (\text{A7})$$

Since our whole calculation has been based on supersimplified rather than conventional cross sections, we use the supersimplified cross sections here also, that is, we use expressions (2.17b) of Rossi and Greisen. This gives for integral values of k

$$M_0(k,0) = \left[1 + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{k} - \frac{2}{k(k+1)} \right]^{-1}. \quad \dots \quad (\text{A8})$$

The "exact" moments given in Table 3 are determined from (A7) and (A8). The fact that f_n for $E \ll E_0$ is independent of E and E_0 shows that in this limit the angular structure function is indeed a function of $(E\theta/E_s)$ only.

It is interesting to observe that these moments do *not* allow a practical determination of the distribution function $v(E_0, E, \theta)$ for small θ . If one takes our best function for v , that is, our v_1 of Section VI, and computes the region of θ which makes the most contribution to the n th moment, this turns out to be well in the "tail" of the function, even for the lowest usable n , that is, $n=2$.

APPENDIX II

The Projected Distribution Function and its Inversion

Let Θ be the angle between the direction of motion of the particle and the shower axis, and let θ be the angle between the projected motion (onto a plane containing the shower axis) and the shower axis. Let θ' be the projected angle for a plane perpendicular to the first one, but also containing the shower axis. Then, in the small angle approximation, we have

$$\Theta = (\theta^2 + \theta'^2)^{1/2}. \quad \dots \quad (\text{B1})$$

Let $f(\theta)$ be the distribution function in the projected angle, and $F(\Theta)$ be the distribution function in the full angle, then clearly

$$f(\theta) = \int_{-\infty}^{+\infty} F[(\theta^2 + \theta'^2)^{\frac{1}{2}}] d\theta' = 2 \int_{\theta}^{\infty} \frac{u F(u)}{(u^2 - \theta^2)^{\frac{1}{2}}} du. \quad \dots \quad (\text{B2})$$

This is an integral equation for $F(u)$ if $f(\theta)$ is known. Since the kernel is a function of the ratio u/θ only, this integral equation can be solved directly by means of Mellin transforms. The details are given by Kalos (1952). The result is

$$F(u) = -(\pi u^2)^{-1} \int_u^{\infty} \frac{x d[xf(x)]}{(x^2 - u^2)^{\frac{1}{2}}} dx. \quad \dots \quad (\text{B3})$$

If $f(x)$ is an exponential e^{-x} , (B3) leads to the Bessel function $(2/\pi)K_0(u)$. The relevant integral representation of $K_\nu(z)$ is formula (4), p. 172 of Watson (1948).

APPENDIX III

The Friedman Variation Principle

This variation principle is contained in a report by Friedman (1949). We would like to thank Dr. F. L. Friedman for permission to publish this slightly modified version of his proof.

The distribution functions used in this work can be considered to form a linear vector space. For example, consider two functions of energy and inclination, $\varphi(E, \theta)$ and $\psi(E, \theta)$. We define their scalar product to be

$$(\varphi, \psi) = \int_0^{\infty} dE \int_{-\infty}^{\infty} d\theta \varphi(E, \theta) \psi(E, \theta). \quad \dots \quad (\text{C1})$$

Functions of x can be treated in the same way; in defining the product for functions of t the integration is limited to the range 0 to ∞ .

All linear operators used may be written in the form of integrals of the type

$$K\varphi(E, \theta) = \int_0^{\infty} dE' \int_{-\infty}^{\infty} d\theta' \varphi(E', \theta') k(E', E, \theta - \theta'). \quad \dots \quad (\text{C2})$$

As a rule the kernel k vanishes when E exceeds E' . The unit operator, I , defined by the equation $I\varphi = \varphi$, is associated with the kernel $\delta(E - E')\delta(\theta - \theta')$.

The effect of two operators of this type applied in turn to a function is given by

$$JK\varphi(E, \theta) =$$

$$\int_0^{\infty} dE'' \int_{-\infty}^{\infty} d\theta'' \int_0^{\infty} dE' \int_{-\infty}^{\infty} d\theta' \varphi(E', \theta') k(E', E'', \theta'' - \theta') j(E'', E, \theta - \theta''), \quad \dots \quad (\text{C3})$$

so that the product of the two operators J and K in that order is an integral operator whose kernel is $Jk(E', E, \theta)$.

Consider the problem of inverting a linear operator. That is, given the non-singular operator A , it is required to find B such that

$$AB = BA = I. \quad \dots \quad (\text{C4})$$

We shall write a variational expression for the matrix elements of the operator B . Suppose that V is an operator, "close to" the inverse operator B , except perhaps for a constant multiplier λ , that is,

$$V = \lambda(B + \varepsilon R), \quad \dots \dots \dots (C5)$$

where ε is by assumption a small quantity.

Let $\varphi_1, \varphi_2, \dots$ be a complete set of base functions in the vector space in question. We define

$$F_{nm} = \frac{(\varphi_n, V \varphi_m)^2}{(\varphi_n, V A V \varphi_m)}. \quad \dots \dots \dots (C6)$$

The array F_{nm} defines an operator in the representation of the base functions φ_n . The operator F defined in this way depends on the choice of base. Since the operators V and A are linear, we may use (C5) to write the last equation as

$$\left. \begin{aligned} F_{nm} &= \frac{(\varphi_n, (B + \varepsilon R) \varphi_m)^2}{\{\varphi_n, (B + \varepsilon R) A (B + \varepsilon R) \varphi_m\}} \\ &= \frac{(\varphi_n, B \varphi_m)^2 + 2\varepsilon(\varphi_n, B \varphi_m)(\varphi_n, R \varphi_m) + (\varphi_n, R \varphi_m)^2 \varepsilon^2}{(\varphi_n, B A B \varphi_m) + \varepsilon(\varphi_n, B A R \varphi_m) + \varepsilon(\varphi_n, R A B \varphi_m) + \varepsilon^2(\varphi_n, R A R \varphi_m)}. \end{aligned} \right\} \dots \dots \dots (C6a)$$

With the use of (C4) the last equation may be written

$$\left. \begin{aligned} F_{nm} &= \frac{(\varphi_n, B \varphi_m)^2 + 2\varepsilon(\varphi_n, B \varphi_m)(\varphi_n, R \varphi_m) + \varepsilon^2(\varphi_n, R \varphi_m)^2}{(\varphi_n, B \varphi_m) + 2\varepsilon(\varphi_n, R \varphi_m) + \varepsilon^2(\varphi_n, R A R \varphi_m)} \\ &= (\varphi_n, B \varphi_m) + \text{terms of order } \varepsilon^2. \end{aligned} \right\} \dots \dots (C6b)$$

Thus the elements of the array F have stationary values when V is near λB . Furthermore when $V = \lambda B$, the elements of F are the elements of B itself, in the representation in which the φ_n are base functions. By finding the stationary values of F_{nm} with respect to variations in V , we get B , that is, the inverse of A . Alternatively, we may use (C6) as an iteration scheme; if V is an approximation to the inverse operator, then the operator F is closer to the true inverse, provided the process converges.

In the special case of the vector space introduced above for the distribution function, an appropriate set of base functions is

$$\varphi_n = \delta(E - E_n) \delta(\theta - \theta_n), \quad \dots \dots \dots (C7)$$

where n stands for the two continuous indices E_n and θ_n . The element $(\varphi_n, K \varphi_m)$ taken with the operator defined in (C2) and two different base functions of this kind is just the kernel $k(E_m, E_n, \theta_n - \theta_m)$. If we use this basis throughout the equation (C6) and apply (C3), we get a variational expression for the kernel $b(E_0, E, \theta)$ of the integral operator $B = A^{-1}$. This expression is

$$f(E_0, E, \theta) = \frac{[v(E_0, E, \theta)]^2}{\int_0^\infty dE' \int_{-\infty}^\infty d\theta' [A v(E_0, E', \theta')] v(E', E, \theta - \theta')}. \quad \dots \dots (C8)$$

Here v is a trial function for b and f is the iterated function.

THE ALBEDO FOR THE ATOMIC SCATTERING OF OPTICAL RADIATION

By R. G. GIOVANELLI* and J. T. JEFFERIES*

[Manuscript received June 10, 1954]

Summary

A general method for evaluating scattering is discussed for a multilevel atom. It is shown that, in the special case when the gas is either opaque or highly transparent to every other spectral line, the effects of interlocking on a given spectral line can be disregarded and the scattering readily computed.

I. INTRODUCTION

The transfer of radiation through an emitting and absorbing medium can be described by the so-called equation of transfer, which in turn is dependent on the monochromatic source function B_ν defined by $B_\nu = E_\nu/\alpha_\nu$, E_ν and α_ν being the emission and absorption coefficients. If emission in a given spectral line arises as a result of previous absorption in the same spectral line, it is referred to as scattering, and is said to be coherent if the re-emission is of the same frequency as that absorbed, non-coherent if not. The albedo for single scattering, denoted by $1 - \lambda$, is defined as the fraction of absorbed radiation which is subsequently scattered.

Thus for isotropic, coherent scattering the source function may be written

$$B_\nu = (1 - \lambda) \frac{J_\nu}{4\pi} + b_\nu, \quad \dots \dots \dots (1.1)$$

where J_ν is the total intensity of the radiation and b_ν is a term dependent on collisional excitation and absorption of radiation in other spectral lines.

For non-coherent scattering

$$B_\nu = \frac{1 - \lambda}{4\pi\alpha_\nu} \int J_{\nu'} \alpha_{\nu'} g_{\nu',\nu} d\nu' + b_\nu, \quad \dots \dots \dots (1.2)$$

where $(1 - \lambda)g_{\nu',\nu}d\nu'd\nu$ is the fraction of energy absorbed in the frequency range ν' to $\nu' + d\nu'$ which is scattered into the range ν to $\nu + d\nu$. In a simple case of non-coherent scattering, the scattered radiation is redistributed proportionally to α_ν , so that

$$g_{\nu',\nu} = \frac{\alpha_\nu}{\int \alpha_\nu d\nu},$$

and so

$$B_\nu = \frac{1 - \lambda}{4\pi} \frac{\int J_\nu \alpha_\nu d\nu}{\int \alpha_\nu d\nu} + b_\nu. \quad \dots \dots \dots (1.3)$$

* Division of Physics, C.S.I.R.O., University Grounds, Sydney.

Another important case occurs in the scattering of continuous radiation which is distributed as $\exp(-h\nu/k\theta)$ for $\nu \geq \nu_0$, the frequency of the beginning of the continuum. Here

$$B_\nu = \frac{1-\lambda}{4\pi\alpha_\nu} \left(\frac{h}{k\theta} \right) \exp \left[\frac{-h(\nu-\nu_0)}{k\theta} \right] \int \alpha_\nu J_\nu d\nu + b_\nu. \quad \dots\dots (1.4)$$

In general, owing to the dependence of b and λ on radiation intensities, simultaneous equations of transfer are involved in describing the radiant intensity and the parameter of importance is not an individual λ —which then has little meaning—but some interlocking parameter obtained by elimination of the other radiation intensities.

Most attention has been devoted to scattering in connexion with the formation of absorption lines by atmospheres in local thermodynamic equilibrium, and a detailed account of techniques applicable has been given by Woolley and Stibbs (1953).*

A method will now be given for simplifying such problems when, for any line other than the one under discussion, the medium is either opaque or transparent. In these cases it will be shown that the quantities b_ν and λ depend only on collisions and external irradiation. As an example of the application of the method general expressions for these two quantities are obtained for the case of an atom with four energy levels.

II. EFFECT OF INTERLOCKING

The equation of secular equilibrium for the population N_j of the j state may be written as

$$\sum_s P_{js} N_j = \sum_s P_{sj} N_s, \quad \dots\dots\dots (2.1)$$

where P_{sj} is the rate of transition $s \rightarrow j$ per atom in the s state, and is the sum of a radiative term A_{sj} (which would be zero for a forbidden transition) and a collision term R_{sj} .

By solving the set of simultaneous equations (2.1), the ratios of the populations may be obtained in terms of the transition rates. Thus in a four-level atom it is found that

$$\frac{N_b}{N_a} = \frac{P_{ab}(1-p_{cd}p_{dc}) + P_{ac}(p_{cb} + p_{cd}p_{db}) + P_{ad}(p_{db} + p_{dc}p_{cb})}{P_{ba}(1-p_{cd}p_{dc}) + P_{bc}(p_{ca} + p_{cd}p_{da}) + P_{bd}(p_{da} + p_{dc}p_{ca})}, \dots (2.2)$$

where

$$p_{rs} = \frac{P_{rs}}{\sum_t P_{rt}}. \quad \dots\dots\dots (2.3)$$

This equation shows how N_b/N_a , and hence the source function for (b,a) radiation, depends on the intensity of each line emitted by the atom.

Now, if the medium is optically thin ($\tau_0 \ll 1$) in any line, self-absorption in this line can be neglected with respect to other excitation processes, and radiative excitation in this line can be due only to irradiation from outside the atmosphere.

* Woolley, R. v. d. R., and Stibbs, D. W. N. (1953).—"The Outer Layers of a Star." Ch. 8. (Oxford Univ. Press.)

Again, if the atmosphere is of sufficient optical depth then

$$A_{rs}N_r = A_{sr}N_s. \quad \dots\dots\dots (2.4)$$

On cancelling these terms in (2.1), except for the transition (a,b) , P_{rs} and P_{sr} are replaced by R_{rs} and R_{sr} respectively, with corresponding changes in (2.2) and (2.3).

If for each spectral line other than the line (a,b) itself, either of the two above conditions applies, then the effects of interlocking on the line (a,b) may be disregarded.

III. DETERMINATION OF λ AND b

In such a case, N_b is the sum of a term \mathbf{N}_b proportional to the rate of absorption of (a,b) radiation, and a term n_b which is explicitly independent of absorption. Then

$$A_{ba}\mathbf{N}_b = (1 - \lambda_{ab})A_{ab}N_a. \quad \dots\dots\dots (3.1)$$

Thus

$$1 - \lambda_{ab} = \frac{A_{ba}(1 - p_{cd}p_{dc})}{D}, \quad \dots\dots\dots (3.2)$$

where D is the denominator on the right-hand side of (2.2).

Hence

$$\lambda_{ab} = \frac{R_{ba}(1 - p_{cd}p_{dc}) + P_{bc}(p_{ca} + p_{cd}p_{da}) + P_{bd}(p_{da} + p_{dc}p_{ca})}{D}. \quad \dots (3.3)$$

The other quantity, b , involved in the source function can also be obtained from (2.2), for

$$b_{ab} = \frac{\rho_{ab}n_b}{4\pi N_a}, \quad \dots\dots\dots (3.4)$$

where ρ_{ab} can be found from expressions for the emission and absorption coefficients. Thus

$$b_{ab} = \frac{\rho_{ab}}{4\pi} \cdot \frac{R_{ab}(1 - p_{cd}p_{dc}) + P_{ac}(p_{cb} + p_{cd}p_{db}) + P_{ad}(p_{db} + p_{dc}p_{cb})}{D}, \quad \dots (3.5)$$

where in both (3.3) and (3.5) for lines of high optical depth P_{rs} is replaced by R_{rs} and for lines of low optical depth self-absorption may be neglected.

IV. CONCLUSIONS

The solution of the transfer equations is simplest in the case of a gas in which the values of R and P in (3.3) and (3.5) can be taken as constant throughout. This will be so in a uniform density isothermal atmosphere of small optical depth in the resonance lines.

Where the optical depth is high in the resonance lines, condition (2.4) breaks down near the boundaries, but this will not usually have a serious effect; the main result will be a reduction in the central intensity of the first resonance line, and a corresponding increase in the central intensity of some of the other lines.

When the atmosphere is of intermediate optical thickness in one or more lines, or near the boundaries of the gas, equation (3.3) still gives formally the effective scattering parameter, though in these cases λ varies with the intensities in the other lines. Unless these can be shown to be negligible, the radiation intensities in such cases are to be found only by solving the equations of transfer simultaneously.

THE EMISSION OF RADIATION FROM MODEL HYDROGEN CHROMOSPHERES. II

By J. T. JEFFERIES* and R. G. GIOVANELLI*

[*Manuscript received August 5, 1954*]

Summary

An improved method is presented for calculating the characteristics of the radiation field of $H\alpha$, $L\alpha$, $L\beta$, and the Lyman continuum emitted by model hydrogen atmospheres which are isothermal at one of a number of kinetic temperatures in the range 10^4 to 2.5×10^5 °K. It is found that earlier estimates of the intensities of these lines need revision and improved values have been obtained.

I. INTRODUCTION

Excitation in and the quantity of radiation emitted by high temperature hydrogen atmospheres have been discussed by a number of authors. Basically, excitation is due to collisions; self-absorption of radiation may, however, cause a profound modification of excitation conditions. Since the radiation emitted in or transmitted through such an atmosphere may undergo scattering or absorption, excitation and the transfer of radiation are interrelated in a manner which is further complicated by possible changes in wavelength upon interaction between radiation and atoms.

The problem may be formulated in terms of a set of simultaneous equilibrium equations, one for the population of each atomic level, and a set of second order differential equations, one for the transfer through the atmosphere of radiation of each wavelength. But the solution of these equations has proved so formidable that substantial simplifications have always been unavoidable. The two most important of these have been

- (i) a restriction in the number of energy levels and corresponding wavelengths considered, and
- (ii) the derivation of approximate rates of emission and absorption by assuming the radiation intensities and then estimating excitation conditions.

Thus, in considering the emission of radiation from model solar chromospheres, Giovanelli (1949) and Jefferies (1953) assumed black body radiation at 5000 °K except in the Lyman lines and continuum, while Thomas (1949) and Matsushima (1952) assumed either no radiation or black body radiation at 6000 °K other than in the Lyman lines.

Nevertheless, assumptions as to radiation intensities, particularly of $H\alpha$, can have a significant effect on the computed emissions, not so much of the

* Division of Physics, C.S.I.R.O., University Grounds, Sydney.

Lyman α line but of the other Lyman and subsidiary lines. We present here a method which substantially avoids these assumptions, the $H\alpha$ intensity being carried as a parameter.

Previous work of the present authors has differed from that of Thomas and of Matsushima in that we have restricted consideration to the states and substates of principal quantum number 1, 2, and 3 and to the ionized state. Thomas considered states of principal quantum number 1–10 and the ionized state, collisions other than with ground state atoms being neglected. Matsushima extended Thomas's results by including collisions with excited atoms.

There is considerable simplification in ignoring the fine structure of the quantum states, and there is now some evidence that this may be justified. Giovanelli (1948) pointed out that transitions between the $2S$ state and the $2P$ state could be ignored unless the appropriate collision cross section was abnormally large. But Purcell (1952) has shown that the effective cross section for collision with positive ions may be as high as $10^7\pi a_0^2$, which is sufficient to ensure that the principal transitions involving the $2S$ state will be to and from the $2P$ state. At electron concentrations of the order of $5 \times 10^{11} \text{ cm}^{-3}$ or more, this alone would ensure that the $2S$ and $2P$ populations are in the ratio of their statistical weights. Further, while computations ignoring $2S \rightleftharpoons 2P$ transitions have shown a strong overpopulation in the $2S$ state in the presence of low radiation densities (i.e. very thin atmospheres), the Lyman radiation intensities in hot atmospheres will normally ensure that the populations of even the $2S$ and $2P$ substates are of the same order. This follows, for example, from explicit solutions of the equilibrium equations given by Giovanelli (1949) and Jefferies (1953). A high $2S \rightleftharpoons 2P$ collision rate will promote an even closer approach to populations distributed according to the statistical weights. We therefore propose to ignore substates and consider the simpler problem so formulated. As before, only the three lower quantum states and the ionized state will be considered.

The present work is confined to a range of physical conditions in which excitation by collisions between neutral atoms may be neglected. This will almost certainly apply for temperatures of about 10^4 °K and higher, and may apply for lower temperatures depending on the cross sections (as yet unknown) for excitation from the excited states.

II. EXCITATION AND COLLISION RATES

The total rate of excitation per unit volume from state j to state l is denoted by $P_{jl}N_j$, where N_j is the population of the j state. P_{jl} may be dissected into a radiative component, A_{jl} , and a collisional component, R_{jl} , which includes the electron concentration N_e as a factor.

The rates of collision excitation between substates have been discussed by Jefferies (1953)—henceforth referred to as paper I. Owing to an error in numerical integration, the collision rates in Table 1 of that paper are too large by a factor of 2. The rates used here are the corrected values.

III. THE EQUATION OF RADIATIVE TRANSFER

With Eddington's approximation, the equation of radiative transfer is

$$\frac{1}{3} \frac{d^2 J}{d\tau^2} = J - 4\pi B, \dots\dots\dots (3.1)$$

which may be written, for a coherently scattering atmosphere, as

$$\frac{1}{3} \frac{d^2 J}{d\tau^2} = \lambda J - 4\pi b, \dots\dots\dots (3.2)$$

where $1 - \lambda$ is the fraction of the absorbed radiation which reappears as scattered radiation in the same line.

The present authors (Giovannelli and Jefferies 1954) have shown that in a four-level atom

$$\lambda_{ab} = \frac{R_{ba}(1 - p_{cd}p_{dc}) + P_{bc}(p_{ca} + p_{cd}p_{da}) + P_{bd}(p_{da} + p_{dc}p_{ca})}{P_{ba}(1 - p_{cd}p_{dc}) + P_{bc}(p_{ca} + p_{cd}p_{da}) + P_{bd}(p_{da} + p_{dc}p_{ca})}, \dots (3.3)$$

where P_{lj} is the rate of transition $l \rightarrow j$ per atom in the l state, and is the sum of a radiative term A_{lj} and a collision term R_{lj} ; and

$$p_{lj} = \frac{P_{lj}}{\sum_k P_{lk}}. \dots\dots\dots (3.4)$$

Further,

$$4\pi b_{ab} = \rho_{ab} \frac{R_{ab}(1 - p_{cd}p_{dc}) + P_{ac}(p_{cb} + p_{cd}p_{db}) + P_{ad}(p_{db} + p_{dc}p_{cb})}{D}, \dots (3.5)$$

where D is the denominator of (3.3), and

$$\rho_{ab} = \frac{8\pi h^3 \nu_{ab}^3}{c^2} \frac{\omega_a}{\omega_b}, \dots\dots\dots (3.6)$$

ω_a and ω_b being the statistical weights of the states, for line transitions. For transitions involving the continuum

$$\rho_{ab} = \frac{\pi h^4 \nu_{ab}^3}{c^2} \left(\frac{2}{\pi m k T} \right)^{3/2} N_e. \dots\dots\dots (3.7)$$

For transitions for which the atmosphere is of great optical depth, P_{ij} and P_{ji} may be replaced by R_{ij} and R_{ji} , while for transitions for which the atmosphere is of small optical depth, self-absorption may be neglected; and, if either of these conditions applies to every spectral line other than the one under consideration, interlocking can be neglected.

In an atmosphere in which λ and b do not vary with τ , the solution of (3.2) may be written

$$J = \frac{4\pi b}{\lambda} + \alpha \exp(\sqrt{3\lambda}\tau) + \beta \exp(-\sqrt{3\lambda}\tau), \dots\dots\dots (3.8)$$

where the integration constants α and β are determined by the boundary conditions. In particular, for a very thick atmosphere with no external illumination, the emergent intensity is given by

$$J = \frac{4\pi b}{\lambda} \frac{2\sqrt{\lambda/3}}{1 + 2\sqrt{\lambda/3}}, \dots\dots\dots (3.9)$$

while for a very thin atmosphere, $\sqrt{3\lambda}\tau^1 \ll 1$,

$$J = 4\pi b\tau^1 \text{ or } 8\pi b\tau^1, \quad \dots\dots\dots (3.10)$$

depending on whether reflection at the base is negligible or complete. The symbol τ^1 represents the total optical depth of the atmosphere.

IV. THE GROUND STATE POPULATION

The intensity of any spectral line is closely associated with the populations of the upper and lower states involved, except in the special case of an optically thin atmosphere. Consequently the population of both the ground and ionized states will generally be related to the intensity of the Lyman continuum.

If Lyman line and continuum radiation is negligible, i.e. if the atmosphere is optically very thin, particularly in $L\alpha$, almost all ionizations occur directly by electron collision from the ground state. These are balanced by recombinations, so that

$$\frac{N_1}{N_4} = \frac{\sum P_{4l}}{R_{14}}, \quad \dots\dots\dots (4.1)$$

where N_4 ($=N_e$) is the ion concentration, the subscript denoting the ionized state, and P_{4l} the rate of recombination per ion to the l -quantum state.

In an atmosphere optically thick in the Lyman lines, any excited atom has a high probability of returning direct or by cascade to the ground state, emitting a Lyman quantum. But, except in the unimportant case near the top of the atmosphere, this quantum will be reabsorbed and the process repeated until an ionization occurs. Consequently almost every collision excitation from the ground state results eventually in an ionization. These, together with direct ionizations from the ground state, are balanced by recombinations direct to the ground state, since any recombination to an excited state also results in a sequence of emissions and absorptions of Lyman quanta until an ionization occurs. Thus

$$\frac{N_1}{N_4} = \frac{P_{41}}{R_{12} + R_{13} + P_{14}}. \quad \dots\dots\dots (4.2)$$

This relation is valid provided the rate of superelastic collision from any excited state (effectively the 2-quantum state) is small compared with the rate of ionization from that state. It follows from the values of the excitation rates that this holds for all values of N_e at kinetic temperatures T of 1.5×10^4 °K or higher, and for $N_e \leq 10^{12}$ cm $^{-3}$ for $T = 10^4$ °K.

Provided the atmosphere is optically thin and not strongly irradiated in the Lyman continuum, P_{14} may be replaced by R_{14} in (4.2) so that

$$\frac{N_1}{N_4} = \frac{A_{41}}{\sum R_{1l}}, \quad \dots\dots\dots (4.3)$$

where, since recombinations are predominantly radiative, P_{41} has been replaced by the rate of radiative recombination A_{41} . In (4.1) and (4.3) the ratio N_1/N_4 is independent of electron concentration and of thickness of the atmosphere.

If the atmosphere is not optically thin in the Lyman continuum, we require the rate of absorption A_{14} . The rate of emission of Lyman continuous quanta of frequency ν is given by the expression

$$[4\pi b + (1 - \lambda)J] \alpha_\nu d\nu / h\nu, \quad \dots \quad (4.4)$$

where α_ν is the absorption coefficient. With J written as

$$J = 4\pi \frac{b}{\lambda} \cdot \varphi(\lambda, \tau), \quad \dots \quad (4.5)$$

(4.4) becomes

$$4\pi b \left[1 + \frac{1 - \lambda}{\lambda} \varphi(\lambda, \tau) \right] \frac{\alpha_\nu d\nu}{h\nu}. \quad \dots \quad (4.6)$$

Since the ratio of the rate of emission of such quanta (which equals the rate of ionization in accordance with (4.2)) to the rate of absorption is increased by the factor

$$1 + \frac{1 - \lambda}{\lambda} \varphi(\lambda, \tau)$$

over the corresponding case for $\tau \ll 1$, it follows that

$$\frac{N_1}{N_4} = \frac{A_{41}}{\Sigma R_{1l} \left[1 + \frac{1 - \lambda}{\lambda} \varphi(\lambda, \tau) \right]}, \quad \dots \quad (4.7)$$

which varies between the limit $A_{41}/\Sigma R_{1l}$ and $\lambda A_{41}/\Sigma R_{1l}$ according as $\tau^1 \ll 1$ ($\varphi = 0$) and $\tau, \tau^1 \gg 1$ ($\varphi = 1$).

From (3.3) it follows on neglecting small terms that

$$\lambda_{14} = \frac{R_{41} + \alpha(P_{42} + P_{43})}{P_{41} + \alpha(P_{42} + P_{43})}, \quad \dots \quad (4.8)$$

where $\alpha = (r_{21} + r_{31}) / (p_{24} + p_{34})$ and $r_{ij} = R_{ij} / \Sigma P_{lk}$.

In the range of N_e and T of interest here, the transition rates are such that $\alpha \simeq r_{21}/p_{24}$ and R_{41} is negligible, so that

$$\lambda_{14} \simeq \frac{\alpha(P_{42} + P_{43})}{P_{41} + \alpha(P_{42} + P_{43})}, \quad \dots \quad (4.9)$$

which is approximately proportional to N_e , and so is dependent on τ_{14} . We shall consider only two broad regions of the atmosphere such that, at the head of the Lyman continuum, $\sqrt{3\lambda}\tau_{14}$ is greater than or less than unity. In the lower region

$$\frac{N_1}{N_4} = \frac{A_{41}}{\Sigma R_{1l}} \lambda_{14}, \quad \dots \quad (4.10)$$

with λ_{14} given by (4.9). In the upper region, the intensity of the Lyman continuum will be roughly that coming from a uniform atmosphere with physical conditions the same as those at $\sqrt{3\lambda}\tau_{14} \simeq 1$. It follows from (3.9) and (4.5) that

$$\varphi(\lambda, \tau) \simeq \frac{2\sqrt{\lambda/3}}{1 + 2\sqrt{\lambda/3}}$$

in this region, λ being computed for conditions applying at $\sqrt{3\lambda}\tau_{14}\simeq 1$, so that for λ not too large, (4.7) reduces to

$$\frac{N_1}{N_4} = \frac{A_{41}}{\Sigma R_{1l}} \frac{\sqrt{3\lambda}}{2}, \quad \dots\dots\dots (4.11)$$

where A_{41} and R_{1l} are values at the electron concentration under consideration ($A_{41}/\Sigma R_{1l}$ is actually independent of N_e).

From (4.9) and (4.11) we may compute the values of N_e at $\sqrt{3\lambda}\tau_{14}=1$ and so estimate the electron concentrations for which (4.10) and (4.11) apply. Results are given in Table 1. In the intermediate region where $\sqrt{3\lambda}\tau_{14}\simeq 1$, the ratio N_1/N_4 changes rapidly with depth by a factor of the order of 10.

TABLE 1
VALUES OF N_e AND N_1/N_4 AT VARIOUS TEMPERATURES

T (10^4 °K)	$\frac{N_e}{\sqrt{3\lambda}\tau_{14}=1}$ (cm^{-3})	Corresponding Values of N_1/N_4
1.0	5×10^{10}	$\frac{A_{41}}{\Sigma R_{1l}} \times 1.9 \times 10^{-13} N_e$ ($10^{12} \geq N_e > 5 \times 10^{10}$) $\frac{A_{41}}{\Sigma R_{1l}} \times 9.5 \times 10^{-2}$ ($N_e \ll 5 \times 10^{10}$)
1.5	5×10^{11}	$\frac{A_{41}}{\Sigma R_{1l}} \times \frac{2.2 \times 10^{-13} N_e}{1 + 1.5 \times 10^{-12} N_e}$ ($10^{12} \geq N_e > 5 \times 10^{11}$) $\frac{A_{41}}{\Sigma R_{1l}} \times 2.8 \times 10^{-1}$ ($N_e \ll 5 \times 10^{11}$)
2.5	$> 10^{12}$	—

The computed values of N_1 are shown in Table 2, for a variety of physical conditions.

TABLE 2
VALUES OF N_1/N_e FOR VARIOUS ATMOSPHERIC MODELS

T (°K)	Thin Atmospheres Eqn. (4.1)	Thick Atmospheres Eqn. (4.3)	Adopted
1.0×10^4	5.3×10	8.0×10^{-1}	} See Table 1 2.2×10^{-4} 3.8×10^{-5} 4.4×10^{-6} 7.8×10^{-7}
1.5×10^4	1.8×10^{-1}	9.7×10^{-3}	
2.5×10^4	1.7×10^{-3}	2.2×10^{-4}	
5.0×10^4	3.8×10^{-5}	9.4×10^{-6}	
1.0×10^5	4.4×10^{-6}	1.4×10^{-6}	
2.5×10^5	7.8×10^{-7}	3.1×10^{-7}	

For $T=5 \times 10^4$ °K, the $L\alpha$ and $L\beta$ optical depths are about unity, which makes computations of radiation intensities difficult. In general, we shall interpolate in this range.

V. THE LYMAN CONTINUUM INTENSITY

The emergent intensity at the Lyman series limit follows from the solution of equation (3.2) with λ_{14} given by (4.9) and, provided the optical thickness in $L\alpha$ and $L\beta$ is large,

$$4\pi b_{14} = \rho_{14} \frac{\Sigma R_{1l}}{A_{41}}, \quad \dots\dots\dots (5.1)$$

where ρ_{14} is given by (3.7).

Evaluating λ_{14} and $4\pi b_{14}$ for the electron concentration at $\sqrt{3\lambda}\tau_{14}=1$ and substituting in the solution of (3.2) we obtain the emergent intensity. For kinetic temperatures greater than 1.5×10^4 °K, the atmosphere is thin in the Lyman continuum if $N_0 < 10^{12}$ cm $^{-3}$, and assuming λ and b to be independent of τ , the emergent radiation is, from (3.10),

$$J_c = 4\pi b_{14} \tau_{14} \\ \simeq \rho_{14} N_e z \alpha(1, \nu_0),$$

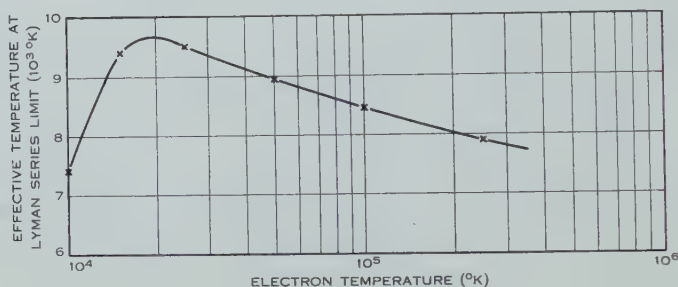


Fig. 1.—Effective temperature at the beginning of the Lyman continuum. $N_0 = 5 \times 10^{11}$ cm $^{-3}$, $\beta = 6 \times 10^{-9}$ cm $^{-1}$.

where z is the scale height for electrons, $\alpha(1, \nu_0)$ the Lyman continuum absorption coefficient per 1S atom and $\rho_{14} N_e$ has the value appropriate to the base of the atmosphere.

Neglecting reflection at the base, the emergent L_c intensity, expressed in terms of the equivalent black body temperature for hemispherical emission, is shown in Figure 1, for $N_0 = 5 \times 10^{11}$ cm $^{-3}$ and $z = 1.66 \times 10^8$ cm. A comparison with the results in paper I shows that, for $T = 10^4$ and 1.5×10^4 °K the emission found here is much greater. This difference arises from neglect of scattering, resulting in too high a ground state population, in the previous work. For the higher temperatures the present results are in agreement with those before.

VI. THE LYMAN LINE INTENSITIES

To evaluate the Lyman line intensities the scattering parameters and source functions are required. From (3.5) we find, on neglecting small terms, that, when the atmosphere is opaque to $L\beta$,

$$4\pi b_{12} \simeq \rho_{12} \frac{\Sigma R_{1l} - p_{41} R_{14}}{A_{21}}, \quad \dots\dots\dots (6.1)$$

and

$$\lambda_{12} \simeq \frac{R_{21} + (P_{24} + P_{23}p_{34})p_{41}}{A_{21}}. \quad \dots\dots\dots (6.2)$$

If the atmosphere is optically thin in $L\alpha$ and $L\beta$ the emergent $L\alpha$ intensity is independent of λ_{12} , while

$$4\pi b_{12} \simeq \rho_{12} \frac{R_{12} + R_{14}p_{42} + (R_{13} + R_{14}p_{43})p_{32}}{A_{21}}. \quad \dots\dots\dots (6.3)$$

From equations (6.1), (6.2), and (6.3), $4\pi b_{12}$ and λ_{12} depend on N_e and so on τ_{12} . However, for atmospheres of high optical depth, the emergent $L\alpha$ radiation is approximately that from a uniform atmosphere of N_e and λ equal to that at $\sqrt{3\lambda}\tau_{12}=1$. For $\sqrt{3\lambda}\tau_{12}^1 < 1$, the emission has been computed for an atmosphere of constant b_{12} using the electron concentration at the base of the atmosphere and assuming complete reflection there.

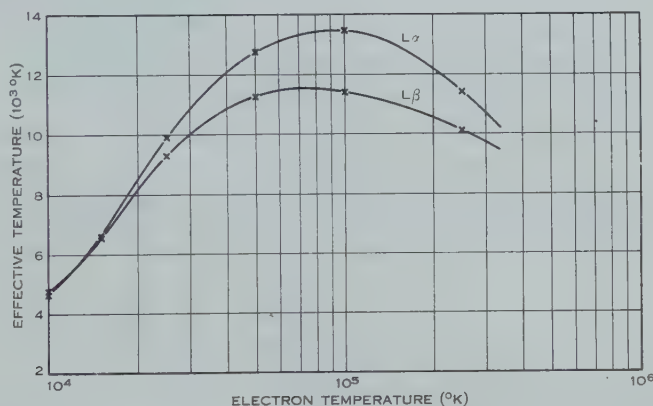


Fig. 2.—Effective temperature at the centre of $L\alpha$ and $L\beta$.
 $N_0 = 5 \times 10^{11} \text{ cm}^{-3}$, $\beta = 6 \times 10^{-9} \text{ cm}^{-1}$.

For $L\beta$, in cases where $\sqrt{3\lambda}\tau_{12} > 1$, neglect of small terms yields

$$4\pi b_{13} \simeq \rho_{13} \frac{\sum R_{1l} - p_{41}R_{14}}{A_{31}}, \quad \dots\dots\dots (6.4)$$

$$\lambda_{13} \simeq \frac{R_{31} + (P_{34} + P_{32}p_{24})p_{41}}{A_{31}}, \quad \dots\dots\dots (6.5)$$

while, if $\sqrt{3\lambda}\tau_{12}^1 \ll 1$,

$$4\pi b_{13} \simeq \rho_{13} \frac{R_{13} + R_{14}p_{43} + (R_{12} + R_{14}p_{42})p_{23}}{A_{31} + A_{32}}. \quad \dots\dots\dots (6.6)$$

The values of λ_{13} and $4\pi b_{13}$ depend on the rate of absorption of $H\alpha$, and so on the $H\alpha$ intensity. Since the mean level of origin for the escaping $L\beta$ radiation is given by $\sqrt{3\lambda}\tau_{13}=1$, a level where the optical depth in $H\alpha$ is small, P_{23} has been computed adopting $H\alpha$ radiation intensities appropriate to the top of the atmosphere (see Section VII).

With boundary conditions that there is no incident radiation on the top of the atmosphere and that perfect reflection occurs at the base, the central intensities of the Lyman lines can now be found, and are shown in Figure 2 in

terms of equivalent hemispherical black body temperatures. These values are not directly comparable with those in paper I, which are calculated for an atmosphere of uniform electron concentration $N_e = 5 \times 10^{11} \text{ cm}^{-3}$ and thickness equal to $1/(6 \times 10^{-9}) \text{ cm}$, while here we have adopted $N_e = N_0 e^{-\beta z}$ with $\beta = 6 \times 10^{-9} \text{ cm}^{-1}$ and $N_0 = 5 \times 10^{11} \text{ cm}^{-3}$. The effect of this difference may be seen from Table 3 where we have given values of N_e such that $\sqrt{3\lambda\tau} = 1$ for $L\alpha$ and $L\beta$. For $T \geq 10^5 \text{ }^\circ\text{K}$, $\sqrt{3\lambda\tau} < 1$.

TABLE 3
ELECTRON CONCENTRATIONS AT THE MEAN LEVEL OF ORIGIN OF THE
ESCAPING $L\alpha$ AND $L\beta$ RADIATION

T ($^\circ\text{K}$)	$\frac{N_e}{\text{at } \sqrt{3\lambda\tau_{12}}=1}$ (cm^{-3})	$\frac{N_e}{\text{at } \sqrt{3\lambda\tau_{13}}=1}$ (cm^{-3})
1.0×10^4	1.4×10^8	2.6×10^7
1.5×10^4	5.2×10^9	1.3×10^9
2.5×10^4	7.1×10^{10}	2.1×10^{10}

VII. THE $H\alpha$ INTENSITY

To find the $H\alpha$ intensity we proceed slightly differently. The intensity of $H\alpha$ is given by the equation

$$\frac{1}{3} \frac{d^2 J}{d\tau^2} = J - \rho_{23} \frac{N_3}{N_2}, \quad \dots\dots\dots (7.1)$$

in which N_3 may be eliminated using the equilibrium equation

$$(P_{31} + P_{32} + P_{34})N_3 = P_{13}N_1 + P_{23}N_2 + P_{43}N_4. \quad \dots\dots\dots (7.2)$$

For regions in which $\sqrt{3\lambda\tau_{13}} \gg 1$, which, for the lower temperatures, is certainly so where $\sqrt{3\lambda\tau_{23}} \simeq 1$, we may put $A_{31}N_3 = A_{13}N_1$, when (7.2) reduces to

$$(R_{31} + P_{32} + P_{34})N_3 = R_{13}N_1 + P_{23}N_2 + P_{43}N_4. \quad \dots\dots\dots (7.3)$$

The transfer equation (7.1) may then be written

$$\frac{1}{3} \frac{d^2 J}{d\tau^2} = J - \rho_{23} \frac{R_{13}N_1 + P_{43}N_4 + P_{23}N_2}{(R_{31} + P_{34} + P_{32})N_2}, \quad \dots\dots\dots (7.4)$$

and so

$$\lambda_{23} \simeq \frac{R_{31} + P_{34} + R_{32}}{A_{32}}. \quad \dots\dots\dots (7.5)$$

Further,

$$4\pi b_{23} \simeq \rho_{23} \frac{R_{13}N_1 + P_{43}N_4 + R_{23}N_2}{A_{32}N_2}, \quad \dots\dots\dots (7.6)$$

a result which may also be obtained from (3.5) on omitting very small terms.

If, for $L\beta$, $\sqrt{3\lambda\tau} \ll 1$, we find

$$\lambda_{23} \simeq \frac{A_{31}}{A_{31} + A_{32}} = 0.56, \quad \dots\dots\dots (7.7)$$

$$4\pi b_{23} \simeq \rho_{23} \frac{R_{23}N_2 + R_{13}N_1 + P_{43}N_4}{(A_{31} + A_{32})N_2}. \quad \dots\dots\dots (7.8)$$

In paper I, λ_{23} was found to be $\sim 1/3$ in the chromosphere and to be almost independent of N_e and T , whereas here, for $\sqrt{3\lambda}\tau_{13} \gg 1$ it is found to be much smaller ($\sim 10^{-2}$) and to vary with the physical conditions. This difference is due to the explicit neglect, in paper I, of $H\alpha$ scattering following the transition $2S \rightarrow 3P \rightarrow 1S$. For $\sqrt{3\lambda}\tau_{13}^1 \ll 1$, λ_{23} is of the same order as that found previously, as expected.

The value of N_2 needed in (7.6) and (7.8) is found from the relationship, valid if the $L\alpha$ optical depth is not too small,

$$\frac{N_2}{N_1} \simeq \frac{A_{12}}{A_{21}} \simeq \frac{J_{12}}{\rho_{12}}, \quad \dots \dots \dots (7.9)$$

while N_1/N_4 is given in Table 2.

As before, the atmosphere is taken as uniform, λ_{23} and $4\pi b_{23}$ being computed for the electron concentration at $\sqrt{3\lambda}\tau_{23} = 1$. If $\sqrt{3\lambda}\tau_{23}^1 \ll 1$, the corresponding quantities adopted are those applicable at the base.

Since in $H\alpha$ the chromosphere is superimposed on a surface of comparable brightness, transmitted photospheric radiation has also to be taken into account.

The solution of (3.2) for the emergent intensity of $H\alpha$ is in such a case (paper I equation (18))

$$J_c = \frac{4\pi b}{\lambda} \left[1 - \frac{\sqrt{\lambda} + \sqrt{\lambda'} + 4\sqrt{\frac{\lambda\lambda'}{3}} \exp(-\sqrt{3\lambda}\tau^1) + (\sqrt{\lambda} - \sqrt{\lambda'}) \exp(-2\sqrt{3\lambda}\tau^1)}{D} \right] \\ + \frac{2(1 + 2/\sqrt{3})\sqrt{\lambda\lambda'} \exp(-\sqrt{3\lambda}\tau^1) J_w}{D}, \quad \dots \dots \dots (7.10)$$

where

$$D = (1 + 2\sqrt{\lambda/3})(\sqrt{\lambda} + \sqrt{\lambda'}) + (1 - 2\sqrt{\lambda/3})(\sqrt{\lambda} - \sqrt{\lambda'}) \exp(-2\sqrt{3\lambda}\tau^1)$$

and J_w is the intensity of the neighbouring continuum.

In this equation, τ^1 is the optical thickness of the chromosphere while λ' represents the photospheric scattering parameter defined by

$$\left. \begin{aligned} 1 - \lambda' = & \frac{\sigma_v}{\sigma_v + \kappa_v + \kappa_0} \\ & - \frac{(\sigma_v + \kappa_v)\lambda_L + \kappa_0}{\sigma_v + \kappa_v + \kappa_0}, \end{aligned} \right\} \quad \dots \dots \dots (7.11)$$

where σ_v and $\sigma_v + \kappa_v$ are respectively the scattering and absorption coefficients, κ_0 is the continuous absorption coefficient of the negative hydrogen ion, and $\lambda_L = \kappa_v/(\sigma_v + \kappa_v)$ is the line scattering parameter of $H\alpha$.

In paper I, λ_L was found from observation to have a value $\sim 1/3$ in the wings of the line (i.e. in the deeper photospheric layers). To calculate λ_L for the photosphere the rates of collision excitation by neutral atoms as well as by electrons are required. Bates and Griffing (1953) have shown that $1S \rightarrow$ excited state transitions in hydrogen caused by H or H^+ impact are of the near adiabatic type with consequent very low cross section at low energies. If similar results

apply for excitation from the 2-quantum state it is justifiable to neglect excitation by neutral atom collisions compared with that by electrons under photospheric conditions. In any case such neglect will give a lower limit to λ_L . From (7.5) it readily follows by considering only electron collisions that, for an atmosphere of kinetic temperature 5000 °K,

$$\lambda_L = \frac{2.3 \times 10^5 + 4.2 \times 10^{-7} N_e}{4.4 \times 10^7 + 4.2 \times 10^{-7} N_e} \dots\dots\dots (7.12)$$

Adopting Munch's (1947) values for the electron pressure in a model solar atmosphere we find that, at $\tau=0.01$ (corresponding to the centre of H α) $\lambda_L=1.7 \times 10^{-2}$; and at $\tau=1.0$ (corresponding to the wings) $\lambda_L=0.25$ in fair agreement with observation, τ being the optical depth in the continuum.

On substituting for the transition rates and putting $\lambda'_{v_0}=\lambda_L=1.7 \times 10^{-2}$, the central intensities of the emergent H α line are obtained as fractions of the adjacent continuum. Results are given in Table 4 for a variety of temperatures and electron concentrations at the base of the atmosphere. For $T=5 \times 10^4$ °K, the results have again been found by interpolation.

TABLE 4
H α CENTRAL INTENSITY
 $\beta=6 \times 10^{-9} \text{ cm}^{-1}$

$N_e \text{ (cm}^{-3}\text{)}$ $T \text{ (}^\circ\text{K)}$	10^{12}	5×10^{11}	2×10^{11}	10^{11}
1.0×10^4	24 (24, 0)*	24 (24, 0)	22 (17, 5)	20 (4, 16)
1.5×10^4	39 (39, 0)	39 (39, 0)	32 (24, 8)	22 (5, 17)
2.5×10^4	60 (60, 0)	60 (59, 1)	42 (32, 10)	26 (7, 19)
5.0×10^4	60 (58, 2)	37 (25, 12)	27 (10, 17)	27 (3, 24)
1.0×10^5	33 (23, 10)	28 (5, 23)	24 (0, 24)	24 (0, 24)
2.5×10^5	26 (2, 24)	24 (0, 24)	24 (0, 24)	24 (0, 24)

* Chromospheric and photospheric contributions, in this order respectively, are included in parentheses after the corresponding H α central intensity. The intensity of the surrounding continuum equals 100 units.

Comparison of these values with the corresponding ones given in paper I shows that the present chromospheric components are in general much greater.

The origin of this difference lies in the adoption in paper I of a definite value for the H α intensity in the earlier stages of the computations. Thus, while the transition rates due to absorption of radiation are in general much greater than collisional rates, the origin of the radiation lies in these collisions together with the radiation incident on the atmosphere. It is not in general legitimate, therefore, to neglect collisional in comparison with radiative transitions; we may approximate between collisional rates or between radiative

rates but not between the two. By approximating to the $H\alpha$ intensity in paper I we effectively did just this and in consequence held down the computed value of the $H\alpha$ intensity.

VIII. DISCUSSION

Comparison of the results of these computations with observations of the emission from the Sun is made rather difficult by lack of knowledge of the chromospheric structure. From the results found in paper I it appeared that, if the $H\alpha$ emitting regions had a temperature in the range discussed, it would lie between 1.5×10^4 and 3.5×10^4 °K. From the present results, however, it would seem that this temperature is less than 1.5×10^4 °K, although the non-uniformity of the chromosphere could modify this conclusion considerably.

Until recently, $H\alpha$ intensities were the only ones available for comparison with theory. However, during recent rocket flights, the solar $L\alpha$ line has been photographed and this will provide very valuable data, the more so as there seems less doubt attached to computed $L\alpha$ intensities than to $H\alpha$ and $L\beta$.

At the present stage it appears that the most useful application of the results would be in interpreting observations of prominences and flares—where conditions are more likely to conform to our model of an isothermal atmosphere—and possibly, from limb observations of the chromosphere, in ascertaining the temperature gradient.

IX. REFERENCES

- BATES, D. R., and GRIFFING, G. (1953).—*Proc. Phys. Soc. Lond. A* **66** : 961.
GIOVANELLI, R. G. (1948).—*Aust. J. Sci. Res. A* **1** : 289.
GIOVANELLI, R. G. (1949).—*Mon. Not. R. Astr. Soc.* **109** : 298.
GIOVANELLI, R. G., and JEFFERIES, J. T. (1954).—*Aust. J. Phys.* **7** : 570.
JEFFERIES, J. T. (1953).—*Aust. J. Phys.* **6** : 22.
MATSUSHIMA, S. (1952).—*Astrophys. J.* **115** : 544.
MUNCH, G. (1947).—*Astrophys. J.* **106** : 217.
PURCELL, E. M. (1952).—*Astrophys. J.* **116** : 457.
THOMAS, R. N. (1949).—*Astrophys. J.* **109** : 480.

THE MULTIPLE SCATTERING OF PROTONS IN NUCLEAR EMULSIONS

By J. R. BIRD* and K. C. HINES*

[Manuscript received September 6, 1954]

Summary

The multiple scattering theories of Williams and Molière have been adapted to give the r.m.s. lateral deflection of protons which lose all their energy in nuclear emulsions. Measurements of 1–5 MeV proton tracks show significant differences from the former theory at low energies and from the latter at higher energies. The introduction of alternative expressions for the minimum angle due to screening does not give a satisfactory explanation of the observed results. It is found, however, that the experimental r.m.s. deflections display the same dependence on maximum single scattering angle as is calculated.

I. INTRODUCTION

During the last few years considerable prominence has been given to the multiple scattering theory of Molière (1948). Some authors have obtained results which appear to indicate that Molière's theory is in better agreement with experiment than, for example, the older theory of Williams (1939). At the present time, however, the issue is somewhat confused. This would seem to be largely due to the inherent uncertainties in the experimental measurements and to variations in procedure adopted for the comparison of theory with experiment.

The two main assumptions which underlie the development of multiple scattering theories are, firstly, that the energy of the particles has not changed over the path in the scatterer and, secondly, that the single scattering angles which contribute to the calculated multiple scattering distribution are small enough for $\sin \Theta$ to be replaced by Θ .

In all experiments performed to date the use of thin scattering foils has been necessary to reduce energy loss but because of this the scattering parameter will be small and there will be considerable statistical uncertainty in the measurements. It seemed desirable, therefore, to find some way of overcoming the first assumption of small energy loss.

The second assumption means that if one wishes to consider particles over an appreciable part of their range there is a limitation to heavy particles. This follows since electrons would be scattered through such large angles that the small angle approximation would not be applicable. Not only this, but for larger angles the penetration depth in the scatterer would not be a satisfactory representation of the path length, as it is taken to be in all the current theories.

* Physics Department, University of Melbourne.

Simple calculation shows that for protons with energies of the order of 5 MeV, which lose the whole of their energy in photographic emulsions, conditions will be such that (with certain reservations) the small angle approximation will be valid up to the end of the range of the particles. This is not true for tracks which exhibit a sudden change in direction attributable to a single scattering event. It is found that the value obtained for the mean of the multiple scattering distribution depends on the upper limit of single scattering angles included in the determination of the mean. This indicates that to obtain a satisfactory basis for comparison the same range of angles must be included in both experimental and theoretical work.

A final consideration is that measurement of length is inherently more accurate than measurement of angle. For this reason it was felt that if the theoretical distribution of the lateral deflection for protons at the end of their range could be obtained, taking account of energy loss and adopting a satisfactory procedure for dealing with the effects of the single scatters, an accurate comparison would be possible with the experimentally observed results. An account of such a comparison is given in the present paper.

II. CALCULATIONS

For protons the spin dependent term in the Mott formula for single elastic scattering is negligible if the energies are less than 10 MeV. If the small angle approximation holds then $\sin \Theta$ may be replaced by Θ and this expression may be written

$$\xi(\Theta)d\omega = 4N \left(\frac{Z^2}{A} \right) r_0^2 \frac{m^2}{p^2 \beta^2} \frac{d\omega}{\Theta^4}, \quad \dots \dots \dots (1)$$

where N is Avogadro's number, r_0 the classical electronic radius, m the mass of the electron, Z and A are respectively the atomic number and the atomic weight of the scattering material, and for low energy non-relativistic protons $p^2 \beta^2$ may be expressed in terms of the particle energy E by $p^2 \beta^2 = 4E^2$.

Following Williams, upper and lower limits of Θ are introduced by making use of the concepts of finite size of the nucleus and screening of the nuclear field by the atomic electrons. The expressions obtained by Williams for these limits are

$$\Theta_{\min.} = \frac{\lambda Z^{1/3}}{2\pi(137)^2 r_0}, \quad \Theta_{\max.} = \frac{\lambda}{2\pi \times 0.57 r_0 Z^{1/3}}, \quad \dots \dots (2)$$

in which λ is the de Broglie wavelength of the incident particles.

Making use of the above limits and of the fact that the single scattering cross section (1) is effectively zero for angles outside these limits, the mean square angle for multiple scattering is given by Rossi and Greisen (1941):

$$\begin{aligned} \langle \Theta^2 \rangle_{dx} &= dx \int_{\Theta_{\min.}}^{\Theta_{\max.}} \Theta^2 \xi(\Theta) 2\pi \Theta d\Theta \\ &= dx 16\pi N \frac{Z^2}{A} r_0^2 \frac{m^2}{p^2 \beta^2} \ln(181 Z^{-1/3}). \quad \dots \dots (3) \end{aligned}$$

This result may be considerably simplified by introducing a new unit of length, the so-called radiation unit, which eliminates most of the constants in formula (3). The radiation unit is defined as

$$\frac{1}{X_0} = \frac{4}{137} N \left(\frac{Z^2}{A} \right) r_0^2 \ln(181Z^{-1/3}). \quad (4)$$

Introducing (4) into (3) the mean square angle of multiple scattering over a distance dt radiation units is

$$\langle \Theta^2 \rangle_{dt} = (E_s^2/4E^2) dt, \quad (5)$$

in which other constants have been combined into the factor E_s^2 in which E_s has the dimensions of energy. Equation (5) is the well-known Rossi-Greisen relation for the mean square angle. Over a finite range of t for which the energy may not be regarded as constant the mean square angle will be

$$\langle \Theta^2 \rangle_t = (E_s^2/4) \int_{E_1}^{E_2} \frac{1}{-(dE/dt)} \frac{dE}{E^2}, \quad (6)$$

where E_1 and E_2 are the energies at 0 and t respectively.

In the calculation of the radiation unit for photographic emulsion it must be remembered that the definition (4) represents the constants in the scattering formula and, in view of the fact that this describes individual processes, a fallacious result will be obtained if mean values of Z and A for the particular compound are substituted into (4). The value obtained in this way for the radiation unit in the C2 emulsion is of the order of $3 \times 10^4 \mu$. Using the correct method and working out a mean value for the quantity $(\ln 181/Z^{1/3})Z^2/A$ for substitution into (4) a value of $7.09 \times 10^4 \mu$ is obtained.

Equation (6) makes it possible to determine the value of the mean square angle for multiple scattering when the energy of the particles does not remain constant over the path in the scattering material. In order to obtain an explicit value some form of range-energy relation must be inserted into the formula. The usual empirical relation used in emulsion work is

$$E = \gamma R^\delta, \quad (7)$$

from which by differentiation there results

$$\frac{dE}{dR} = \alpha E^\beta, \quad (8)$$

where the constants have the values $\gamma = 0.262$, $\delta = 0.575$, $\alpha = 0.0558$, and $\beta = -0.74$.

If the effect of the multiple scattering on the range is neglected, then, denoting the residual range in radiation units by s and the total range by S , it is clear that

$$t' + s = S,$$

where t' is the penetration depth in the scatterer. If n is the number of microns per radiation unit, (7) becomes

$$E = \gamma \{n(S - t')\}^\delta. \quad (7a)$$

Using these results (6) may be integrated to yield

$$\Theta_{r.m.s.} = \sqrt{\langle \Theta^2 \rangle} = \frac{E_s}{2} \frac{(\gamma n \delta)^{-(1+\beta)/2} S^{-\delta(1+\beta)/2} \left[\left(1 - \frac{t'}{S} \right)^{-\delta(1+\beta)} - 1 \right]^{\frac{1}{2}}}{\{(1+\beta)n\alpha\}^{\frac{1}{2}}} \quad \dots (9)$$

A formula having the same form as this equation has been derived by Wilson (1947) but is not as satisfactory for two reasons. Firstly, the empirical range-energy relation is more accurate for energies above 2 MeV than the approximation to the Bethe-Bloch formula used by Wilson and, secondly, this author, in his derivation, has made an approximation which, according to his paper, introduces an error of about 5 per cent.

The relation (7) is satisfactory above 3 MeV, but for lower energies the experimental range-energy results cannot be fitted using a constant γ . For this reason (9) is valid only for particles with energy high enough so that the part of the range during which the energy falls from 2 MeV to zero is small compared with the total range. If this condition is not satisfied then the integration of (6) must be done numerically using empirical range-energy results.

As has been stated in the introduction, comparison with the experimental values is possible only if, instead of taking a theoretical upper limit for single scattering angles, values are chosen to conform with the experimental procedure. For this reason the lateral deflection has been calculated using upper limits of 5, 10, 15, 20, and 25°. Since it is more convenient to measure projected quantities rather than space quantities, use is made of the fact that the mean square projected angle is half the mean square space angle.

Under these conditions equation (5) becomes

$$\langle \theta^2 \rangle_{dt} = \frac{dt E_s^2}{16 E^2} \cdot \frac{\ln (\varphi \times 10^4 \sqrt{E}/2 \cdot 03)}{\ln 181 Z^{-1/3}}, \quad \dots (10)$$

where θ is now a projected angle and φ takes on the arbitrary values given above. From this result the root mean square value of the lateral deflection from the original direction is readily obtained

$$y_{r.m.s.} = \int_0^t \tan \theta_{r.m.s.} dt'. \quad \dots (11)$$

The required value of $y_{r.m.s.}$ is found by inserting the root mean square value of θ from (10) into (11). For protons of initial energy 10 MeV or less the values of $y_{r.m.s.}$ must be obtained by numerical integration.

In the evaluation of (11) successively greater values of φ are introduced and from these results a curve of root mean square lateral deflection versus upper limit of single scattering angle may be plotted. Curves of this kind for four values of initial energy are given in Figure 4.

The values of the lateral deflection are found from the Molière theory using the formula for the mean multiple scattering angle after particles have traversed a distance σ g/cm² of scattering material:

$$\theta_m = \frac{22 \cdot 9 Z}{2 \sqrt{\pi E}} \sqrt{\frac{\sigma}{A}} \{ B^{1/2} + 0 \cdot 982 B^{-1/2} - 0 \cdot 117 B^{-3/2} \}, \quad \dots (12)$$

where the parameter B is related to a second parameter Ω_b which is given by

$$\log_{10}\Omega_b = 8.215 + \log_{10} \left[Z^{-2/3} \left(\frac{\sigma}{A} \right) \alpha^2 / (1.13 + 3.76\alpha^2) \right]. \quad \dots (13)$$

The relation between B and Ω_b is

$$\Omega_b = 1.167e^B/B, \quad \dots (14)$$

and

$$\alpha = 2.095/\sqrt{E}$$

for protons in photographic emulsion.

Since the energy is a function of σ these formulae will only be true over an infinitesimal range of σ .

For finite σ we will have

$$\log_{10}\Omega_b = 8.215 + \log_{10} \int_0^\sigma \frac{0.0269 d\sigma'}{1.13E + 16.5}, \quad \dots (15)$$

in which the constants for the particular scattering substance have been inserted. By integration of (15) and the use of (14) the values of B corresponding to a series of values of σ from the beginning to near the end of the range of the protons are obtained. It is then possible to use the relation

$$\theta_m^2 = \int_0^\sigma \frac{(15.87)^2}{E^2} \{ B^{1/2} + 0.982B^{-1/2} - 0.117B^{-3/2} \}^2 d\sigma' \quad \dots (16)$$

to find values of θ_m along the path.

TABLE 1
THEORETICAL AND EXPERIMENTAL VALUES OF LATERAL DEFLECTION

Energy (MeV)	4.87	3.91	2.66	1.35
R.m.s. deflection (15° limit) (μ)				
Experiment	11.0 \pm 0.2	8.0 \pm 0.3	4.38 \pm 0.06	1.70 \pm 0.03
Rossi-Greisen B				
θ_{\min} , eqn. (17)	11.16	7.96	4.80	2.07
Rossi-Greisen C				
θ_{\min} , eqn. (2)	10.78	7.71	4.60	1.96
Rossi-Greisen E				
θ_{\min} , eqn. (18)	10.39	7.41	4.37	1.73
Mean deflection (all tracks) (μ)				
Experiment	8.9 \pm 0.3	6.7 \pm 0.3	3.5 \pm 0.1	1.34 \pm 0.05
Molière	8.19	5.92	3.38	1.40

Now a defect of the Molière theory is that it is impossible to find a mean square scattering angle. In fact the general expression for the k th moment of the distribution becomes infinite for even values of k . It is thus only possible to find mean values of the various quantities using Molière's results, and a comparison with experiment is only possible after expressing the experimental results as means.

Owing to the fact that the Molière expression for the k th moment is obtained by integrating over all angles from 0 to ∞ , and that it is only because of this choice of limits that the complicated expression for the general moment reduces to the simple form (12), the lateral deflection has not been worked out in this case for a series of upper limits on the single scattering angle considered.

Values of the mean lateral deflection are found using equation (11) with $\theta_{r.m.s.}$ replaced by θ_m of (16). Results for the four different proton energies are plotted to give the curve in Figure 5.

All the calculated results have been collected in Table 1.

III. EXPERIMENTAL PROCEDURE

Tracks of monoenergetic protons were obtained by exposing Ilford C2 emulsions to protons from the reactions ${}^9\text{Be}(d,p){}^{10}\text{Be}$, ${}^6\text{Li}(d,p){}^7\text{Li}$, and ${}^{12}\text{C}(d,p){}^{13}\text{C}$. The exposures were made in a camera which has been described by Martin *et al.*

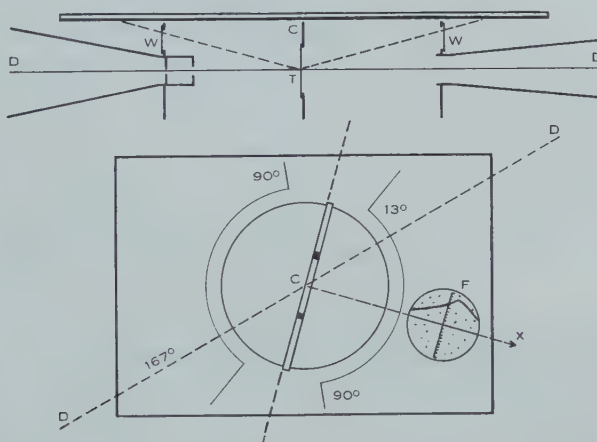


Fig. 1.—Photographic plate showing location in camera and method of predicting initial direction of tracks. *DD*, deuteron beam; *T*, target spot; *W*, “Cellophane” windows; *F*, microscope field of view.

(1949). It contains a photographic plate parallel to the incident deuteron beam and 1 cm above it (Fig. 1). A copper ring supports the plate and the target holder and has “Cellophane” windows which stop the scattered deuterons but allow the more energetic reaction products to reach the outer parts of the plate. Those particles which reach the plate at a distance of 4 cm from the point *C* vertically above the target spot strike the emulsion at an angle of 13° to the plane of the surface. At various positions on one plate, therefore, a range of angles from 13 to 167° relative to the incident deuteron direction is obtainable. This, together with the use of one or more thicknesses of “Cellophane”, provides a range of proton energies from 1 to 5 MeV.

The reactions used were chosen since they have accurately known Q -values and characteristic energy spectra which allow a group of protons with a particular energy to be distinguished without measuring the length of each track. The

use of thin targets and suitable control of the bombarding energy ensures that the reaction products will have a small spread in energy. The length of exposure was chosen in each case to give a high density of tracks in the emulsion. The limiting condition on the track density was provided by the requirement that tracks should not interfere with one another. With a high density the required number of tracks can be obtained in a small area of plate over which the change in energy due to change in angle relative to the direction of the deuteron beam will be small. For the same reason the diameter of the deuteron beam is restricted by diaphragms to 0.020 in.

In order to test the dependence of the width of the scattering distributions on experimental conditions, emulsions from various batches and of various thicknesses were used. They were vacuum dried to constant weight prior to exposure and maintained under vacuum during exposure. A constant value of 3.94 g/c.c. was then assumed for the emulsion density in the calculation of the

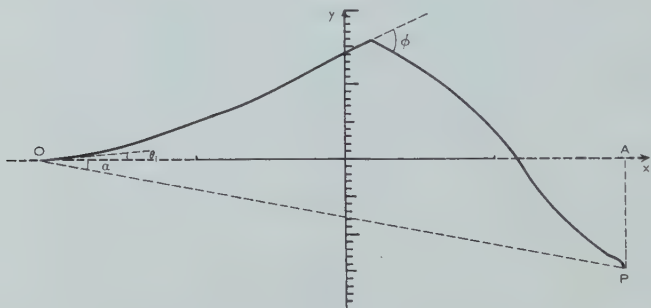


Fig. 2.—Schematic representation of a track and eyepiece graticule.
 θ_i , initial angle; ϕ , single scattering angle; α , average angle,
 x , predicted initial direction; y , lateral deflection.

combined effect of the component elements on the scattering distribution. The processing of the emulsions was varied to some extent, but as far as possible care was taken to prevent distortion and reticulation.

From a knowledge of the relative positions of the deuteron beam, target spot, and photographic plate one may predict the initial direction of tracks which appear on any part of the plate. During exposure the plate rests on the upper edge of the target holder and this is placed accurately at 45° to the direction of the incident beam. The holder prevents scattered deuterons from blackening the plate except where two V-slots occur in the top of the holder at known distances from the point *C* (Fig. 1). In this way the position of the slots is registered on the plate. After processing, therefore, the position of the point *C* and the direction of the target holder can be ascertained.

As a check on the geometry, a mask was constructed to fit in the top of the camera, and marked with the various positions and directions. During loading of the camera a brief exposure to light produces an image of the markings on the surface of the emulsion nearest the glass backing.

The measurements were made with a microscope fitted with a polar stage on which the plates were mounted with the target direction along the 45° line

and with the point *C* over the centre of rotation. The stage can be moved along a line joining the centre of the objective and the centre of the stage. Since the tracks appear to originate at a point above the centre of the stage, their initial direction will coincide with the direction of linear motion, no matter what part of the plate is studied. Location of the plate was checked in each case by observing the behaviour of the image of the mask as the stage moved. In all cases the angular settings of the stage agreed with those of the mask to within $\frac{1}{2}^\circ$. This is consistent with the accuracy possible in the construction of the mask.

In order to measure the lateral deflection of each track from its initial direction use was made of an eyepiece graticule which has a scale and its perpendicular bisector engraved on it (Fig. 2). The orientation of the eyepiece was such that the perpendicular to the scale lay in the direction of linear motion of the stage and was then assumed to indicate the correct initial direction for all tracks. The displacement of the end of each track from this direction may be determined by moving the track so that first the beginning and then the end passes across the scale, and taking the difference between the readings for each position. Each reading was made to the nearest half division, so that the lateral deflection was accurate to the nearest division. The magnification is about 1500 for all tracks, since with the procedure adopted it does not matter if the tracks are more than one field of view in length. With this magnification, one scale division corresponds to less than $1\ \mu$ and is of the order of magnitude of the average grain size, so that no gain in accuracy could be achieved by trying to make measurements more accurately than this.

The actual initial direction of each track, θ_i , could be checked by observing the scale reading at a certain distance from the beginning of the track, chosen so that a deflection of half a division would correspond to an angle of 1° . In this way a rough estimate of the angle was made, allowing for any curvature in this first section of the track. A check of this kind is necessary since an appreciable number of protons are scattered in the "Cellophane" and therefore all particles do not arrive at the emulsion with the same initial direction.

A second eyepiece scale was used in conjunction with an eyepiece protractor to measure the approximate magnitude of any single scattering events which occurred. Reasonable accuracy could be obtained for large angles but, for angles less than 10° , not only was the accuracy of measurement poor but also the chance of detecting such events became increasingly less. No attempt was made to measure angles less than 5° and these were only included when obvious.

A total of about 8000 tracks was measured and the characteristics of each tabulated. If any sign of distortion appeared no tracks were measured in that neighbourhood. Measurements were made at the four energies selected for comparison with the calculations. These values were obtained by measuring the range of about 300 tracks at each energy, and then using the range-energy relation for C2 emulsions. The values found checked satisfactorily with those calculated from the energetics of the particular reaction used and the energy lost in traversing the "Cellophane".

IV. RESULTS

At each proton energy histograms are plotted for the lateral deflections of those tracks which have a correct initial angle, a separate histogram being plotted for each group of tracks having a particular range of magnitude of single scattering events. A typical histogram is shown in Figure 3 and compared with a Gaussian distribution which has the same standard deviation. The characteristic features of a scattering distribution are indicated by the increased number of tracks at the centre and extremes, while at the intermediate deflections

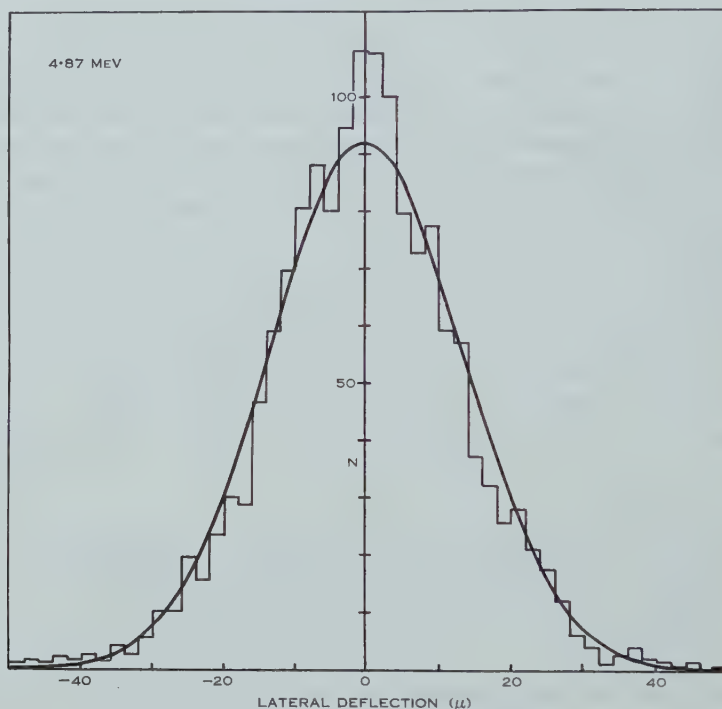


Fig. 3.—Typical histogram of number of tracks *v.* lateral deflection.
Smooth curve is the Gaussian with the same r.m.s. deflection.

there are less than would be found for a Gaussian distribution. The ratio of standard deviation to mean for the experimental curves is found to be larger than the value $\sqrt{\pi/2}$ for a Gaussian distribution, the increase being about 3 per cent. at the higher energies investigated, and nearly 20 per cent. at the lowest energy.

The increase in standard deviation at each energy, as increasing single scattering angles are included, is shown in Figure 4, together with the probable errors calculated from the number of tracks associated with each measurement. A direct comparison can be made between these results and the calculations by the Rossi-Greisen method, but not with the mean values which are obtained from the theory of Molière. The Molière values indicated in Figure 5 are obtained using the observed values of the ratio of the standard deviation to the mean.

The experimental results at the two higher energies are both a combination of two separate measurements, the difference between them being less than the probable errors for each. The results at 3.91 MeV have been analysed in two different ways in order to check the measurements of initial angle. The tracks

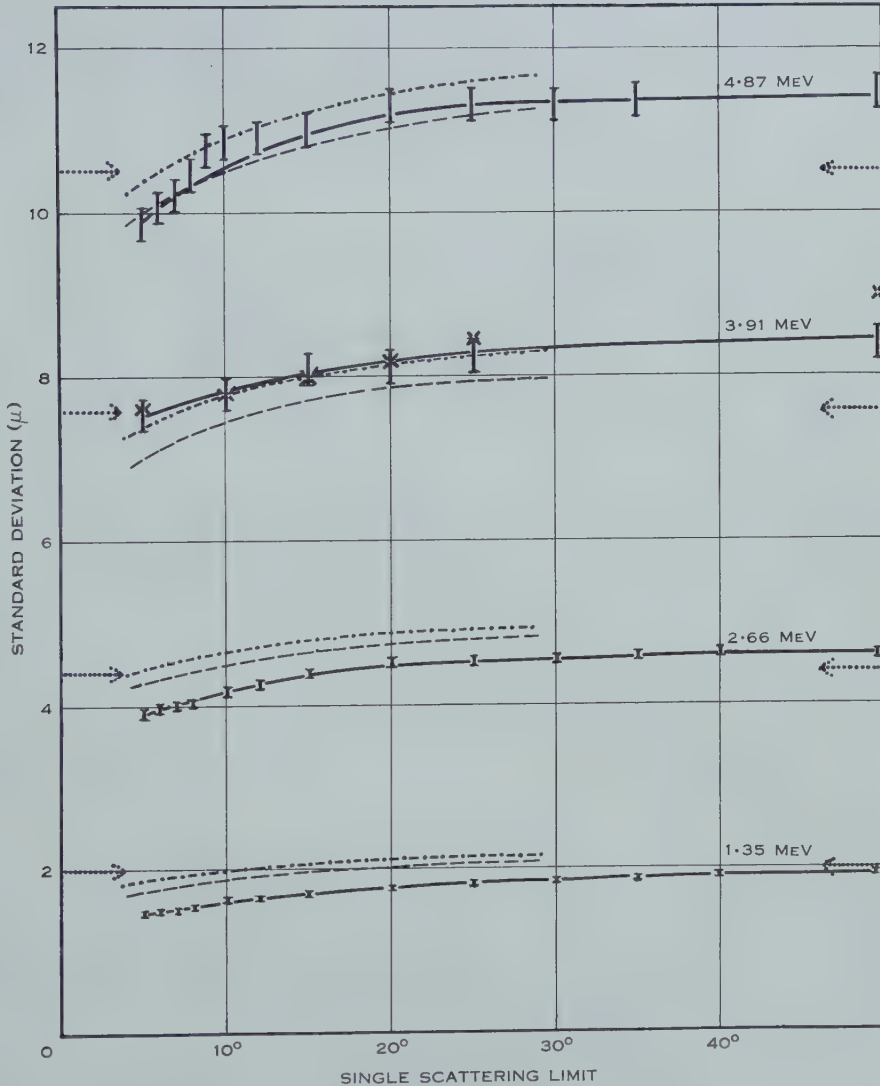


Fig. 4.—Variation of r.m.s. deflection with single scattering limit. Full curves, experimental values; dashed curves, Rossi-Greisen with θ_{\min} from equation (17); dot and dash curves, Rossi-Greisen with θ_{\min} from equation (2); arrows, Molière values.

are divided into groups according to their estimated initial angle and each group analysed separately. The observed centre of each group agrees with that expected from the initial angle, and the standard deviations calculated about these centres are also consistent. The results obtained by combining standard

deviations for all initial angles are plotted in Figure 4 for comparison with the zero-angle results.

The centres of the distributions for zero-angle tracks should be zero, and in all cases the average deflection when divided by the mean range indicates an initial angle of less than $\frac{1}{2}^\circ$. This verifies that the plates have been located during the experiment to this degree of accuracy. Each non-zero average deflection has been used to give a small correction to the standard deviations which should therefore be free from the effect of errors in the geometry of the experiment.

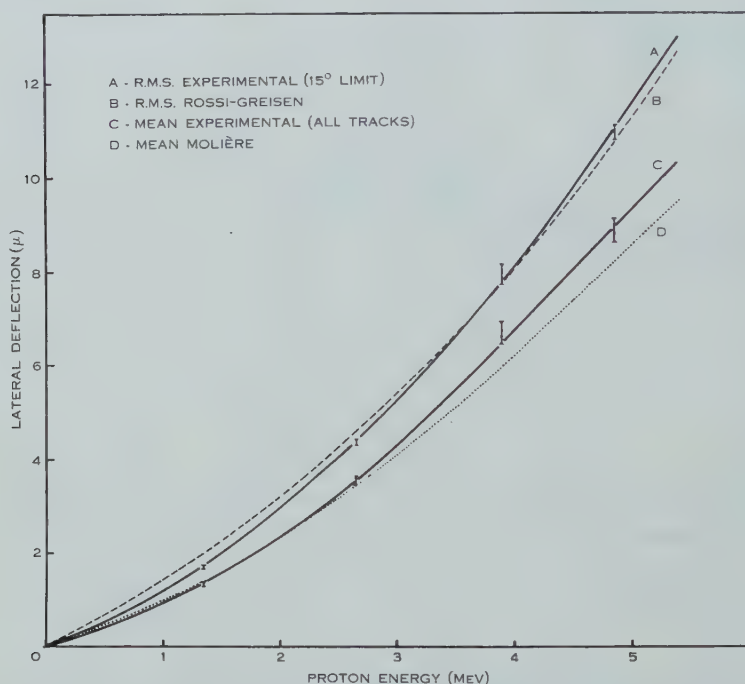


Fig. 5.—Dependence of mean and r.m.s. deflection on proton energy.

The effect of loss of tracks by scattering out of the emulsion has been investigated empirically, since an accurate calculation would be difficult. The number of tracks lost depends on the distance of the point *A* (Fig. 2) from both surfaces of the emulsion and can be obtained by direct observation as well as from the number of tracks having a sufficiently large lateral deflection. The distribution in radial deflection was assumed to be approximately Gaussian with a variance twice that observed for the projected lateral deflection. The tracks having a small range of radial deflection were assumed to be distributed uniformly over an annular strip which may be divided into a number of segments. The projected deflection was plotted for the tracks in each segment, the centre of which lay outside the emulsion, and then summed for all annular strips contributing to the total distribution. The variance obtained in this way for the lost tracks can be used as a correction to the variance of the observed

distribution in lateral deflection. The correction was found to be $\frac{1}{2}$ per cent. for the one plate for which the loss of tracks was considerable, and negligible for the others.

At the lower energies grouping of the tracks into one-division intervals has an appreciable effect on the standard deviation and so Sheppard's correction has been applied (Kendall 1947). The final values for a 15° single scattering limit are included in Table 1, together with the values of the mean deflection obtained by including all events. The dependence of these values on proton energy is illustrated in Figure 5, in which the probable errors used are a combination of the statistical uncertainties and the uncertainty in proton energy.

V. DISCUSSION

The use of arbitrary values of $\theta_{\max.}$ in the calculation of multiple scattering by the Rossi-Greisen method gives results which vary in the same way as the measurements, using the same $\theta_{\max.}$ values (dot and dashed curves in Fig. 4). At small angles the experimental values fall a little faster but this may be due to systematic errors in the detection and measurement of these angles. The two low energy curves are seen to be appreciably higher than the experimental ones.

Since the form of the single scattering expression in equation (1) should be reliable for the protons used, we need to investigate the form of $\theta_{\min.}$ used in the equation of type (3) for projected mean square angle. The dot and dash curves in Figure 4 have been obtained using the value for the projected $\theta_{\min.}$ given in (2). This angle was calculated by Williams on the basis of a simple consideration of the screening of the nuclear field by the atomic electrons. The value found is a direct consequence of the use of the Born approximation in the derivation of the single scattering formula.

On the basis of a more refined consideration of the screening effect Williams has obtained the result

$$\theta_{\min.} = \frac{1 \cdot 75 \lambda Z^{1/3}}{2\pi(137)^2 r_0} \dots\dots\dots (17)$$

In this estimate account is taken of the fact that the cut-off due to screening is not abrupt but is spread over a range of angles. This effective screening angle has been obtained using the calculations of Bullard and Massey (1930) for the Thomas-Fermi field of the atomic electrons.

The dashed curves of Figure 4 have been calculated using this value for $\theta_{\min.}$. They are seen to agree with the experimental results more satisfactorily than the other curves, although the variation with $\theta_{\max.}$ remains unaltered. A comparison (Fig. 5) of the results for a particular value of $\theta_{\max.}$ and using $\theta_{\min.}$ from (17) shows reasonable agreement with experiment above 3.5 MeV, but at lower energies the calculated results are still considerably higher.

A comparison of the experimental results with the theory of Molière, however, indicates the opposite effect (curve *D*). There is reasonable agreement at low energies, but the calculations are low at the higher energies. Both these discrepancies are outside the experimental errors and indicate that neither

approach to the multiple scattering of protons can adequately explain the observed results.

This is indicated more clearly in Figure 6, in which the average angle α (Fig. 2) is plotted against energy. In this figure curves *B* and *C* represent the results for $\theta_{\min.}$ given by (2) and (17) respectively.

Curve *E* has been obtained by working out the lateral deflection again, using the Rossi-Greisen method and the Molière value for $\theta_{\min.}$:

$$\theta_{\min.} = \frac{1.142mZ^{1/3}}{137M\beta} \left\{ 1.13 + \frac{3.76Z^2}{(137)^2\beta^2} \right\}^{1/2}, \quad \dots\dots (18)$$

where M is the mass of the proton.

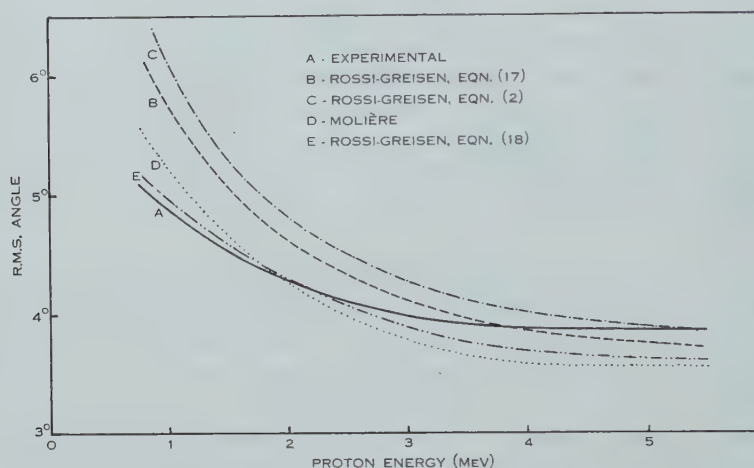


Fig. 6.—Dependence of average angle on energy.

It is interesting to note that this curve lies fairly close to curve *D* found from the Molière theory itself. The slope of these two curves is almost the same although curve *E* is closer to the experimental curve.

The characteristic screening angle does not occur in the Molière theory in the form of a lower limit but is included in the approximate single scattering expression. The dependence of the Molière values on screening angle is thus quite complicated and it is of interest that the values obtained in this way are less satisfactory than those represented by curve *E*. This may be due to the approximations which Molière has been obliged to make.

The shape of the curves in Figure 6 is very similar to those presented by Groetzinger, Berger, and Ribe (1950) and Groetzinger, Humphrey, and Ribe (1952) for the multiple scattering of electrons with energies up to 2 MeV. In both cases the theoretical results are too high at the lowest energies, and decrease too rapidly with increasing energy.

Using 115 MeV electrons Corson (1950) claims better agreement with the theory of Snyder and Scott (1949) than with the Rossi-Greisen theory. Corson uses a value of 2.92 cm for the radiation unit in Ilford G5 emulsions and obtains a value of 0.28°/100 μ for the mean multiple scattering deflection using the

Rossi-Greisen theory. The value of the radiation unit for the emulsion obtained using the correct averaging procedure is 7.41 cm , and gives a mean deflection of $0.18^\circ/100 \mu$. Using the Snyder-Scott theory which does not involve the use of radiation units he finds $0.20^\circ/100 \mu$. In contrast to Corson's conclusion, therefore, his experimental value of $(0.17^\circ \pm 0.02)/100 \mu$ is in better agreement with the Rossi-Greisen result than with that obtained from the more complicated theories of Molière or Snyder and Scott.

Molière's theory, however, is presented in such a way that it is necessary to choose with care the method of evaluation for any particular case. This is illustrated by the work of Spencer and Blanchard (1954). These authors have shown that the Molière distribution may be improved by the introduction of a more accurate single scattering expression for the large angles which fits the multiple scattering distribution for small angles better than the Molière large angle formula. With this modification they obtain a curve for the multiple scattering of 15.7 MeV electrons which differs from the observed shape obtained by Hanson *et al.* (1951). This is in contrast with the claim made by Hanson *et al.* that their results were in good agreement with Molière's.

Other experiments on the scattering of electrons and protons do not show any systematic agreement with either theory, although where the theory of Williams gives wider distributions than the Molière theory the experimental values are usually lower than the former. The differences are seldom found to be greater than 10 to 15 per cent. and often are of the same magnitude as the experimental errors.

VI. CONCLUSIONS

Using conditions under which a simple single scattering law is valid, the calculation of multiple scattering by the Rossi-Greisen method gives a satisfactory dependence on the maximum scattering angle. It is found, however, that none of the usual expressions for the minimum angle gives an energy dependence which will explain the observed results.

Use of the more complicated Molière expression for the screening angle in the Rossi-Greisen calculations gives a better fit than the \sqrt{E} dependence of the Williams expressions but it still has too strong a dependence on energy. Further measurements are required to investigate these effects for particles of higher energy.

VII. ACKNOWLEDGMENTS

Our thanks are due to Professor L. H. Martin and Associate Professor C. B. O. Mohr for their interest in this work.

VIII. REFERENCES

- BULLARD, E. C., and MASSEY, H. S. W. (1930).—*Proc. Camb. Phil. Soc.* **26** : 556.
CORSON, D. R. (1950).—*Phys. Rev.* **80** : 303.
GROETZINGER, G., BERGER, M. J., and RIBE, F. L. (1950).—*Phys. Rev.* **77** : 584.
GROETZINGER, G., HUMPHREY, W., and RIBE, F. L. (1952).—*Phys. Rev.* **85** : 78.
HANSON, A. D., LANZL, L. H., LYMAN, E. M., and SCOTT, M. B. (1951).—*Phys. Rev.* **84** : 634.
KENDALL, M. G. (1947).—"The Advanced Theory of Statistics." (Griffin & Co. Ltd. : London.)
MARTIN, L. H., BOWER, J. C., DUNBAR, D. N. F., and HIRST, F. (1949).—*Aust. J. Sci. Res. A* **2** : 25.

- MOLIÈRE, G. (1948).—*Z. Naturf.* **3a**: 78.
 ROSSI, B., and GREISEN, K. (1941).—*Rev. Mod. Phys.* **13**: 240.
 SNYDER, H. S., and SCOTT, W. T. (1949).—*Phys. Rev.* **76**: 220.
 SPENCER, L. V., and BLANCHARD, C. H. (1954).—*Phys. Rev.* **93**: 114.
 WILLIAMS, E. J. (1939).—*Proc. Roy. Soc. A* **169**: 531.
 WILSON, R. R. (1947).—*Phys. Rev.* **71**: 385.

APPENDIX I

Range Straggling due to Multiple Scattering

A procedure similar to that in Section II may be adopted to obtain an estimate of the effect of multiple scattering on the straggling of protons in the energy range considered. Because of multiple scattering some particles follow more devious tracks than others and the perpendicular penetration depth in the scattering material will not be the same for all. In actual fact the measured range will not be the penetration depth (OA , Fig. 2), but the distance between the point of entry of the particle into the emulsion and the point where the track ends (OP).

The path length corresponding to the root mean square angle will be

$$L = \int_0^t \frac{dt'}{\cos \Theta_{r.m.s.}} \dots\dots\dots (19)$$

If the distance t is small enough so that energy loss may be neglected one may use equation (5) to obtain the result

$$L = t - \left(\frac{E_s}{2E}\right)^2 \frac{t^2}{2 \times 2!} + 5 \left(\frac{E_s}{2E}\right)^4 \frac{t^3}{3 \times 4!} + \dots\dots\dots (20)$$

For most applications, however, this is not a satisfactory approximation and one must use instead the relation (6) to obtain

$$L = \int_0^t \cos^{-1} \left[\frac{E_s}{2} \sqrt{\left\{ \int_{E_s}^{E_1} \frac{1}{-(dE/dt)} \frac{dE}{E^2} \right\}} \right] dt'. \dots\dots (21)$$

This expression has been evaluated for a series of values of initial proton energy. The values of L so found may be regarded as a measure of the mean value of the path length for all particles. The difference between this and the range of the particles will be a convenient measure of the magnitude of the multiple scattering effect on the straggling. Thus it is required to work out $L - S$, where S is the OP of Figure 2. Values of L have been expressed as a percentage of the range and it is found that for protons with energy of the order of 10 MeV the effect is about $\frac{1}{2}$ per cent. For protons with energy 0.5 MeV the value has increased only to 1.18 per cent. In view of the fact that the observed straggling for protons with these energies at the end of their range is about 6 per cent., it is seen that the major contribution to the total straggling must come from the variations in ionization energy loss.

MEASUREMENTS OF THE COSMIC RAY NEUTRON RATE IN THE HIMALAYAS AND AUSTRALIAN ALPS

By K. B. MATHER*

[*Manuscript received August 30, 1954*]

Summary

The variation of α -track density in boron-loaded emulsions and the star rate in C2 emulsions have been measured at the same sites at mountain altitudes at two latitudes. The α -count is proportional to the slow neutron density in the atmosphere while the stars arise chiefly from fast neutrons. The variation with altitude can be represented in each case by a simple exponential of the form: $I = I_0 e^{-x/L}$. Values were derived for the attenuation length L for each latitude:

geomagnetic lat. 21 °N., $L = 148$ (slow neutrons), 139 (stars);

geomagnetic lat. 45 °S., $L = 141$ (slow neutrons), 130 (stars).

The close resemblance and direction of variation of L values for neutron flux and star rates is evidence for their generic connexion. The latitude variation of L is less than observed in the upper half of the atmosphere. L values are also somewhat smaller than measured at higher altitudes, consistent with a degradation of the nucleon component with increasing atmospheric depth. The star rate in the atmosphere has been calculated to be ~ 1.3 stars $\text{cm}^{-2} \text{sec}^{-1}$ at geomagnetic latitude 45 °S.

I. INTRODUCTION

It has been recognized since soon after the discovery of the neutron that the cosmic ray flux through the atmosphere includes a neutron component. Because of its lack of charge (and hence immunity from accelerating mechanisms) together with its instability it cannot constitute an appreciable part of the primary radiation. However, it is plausible that neutrons comparable in energy to the primary radiation could be produced as disintegration and exchange products of high energy primaries interacting with oxygen and nitrogen nuclei in the top layers of the atmosphere. Interactions of this kind are observed as large stars in photographic emulsions exposed by balloons at $\sim 100,000$ ft and it seems reasonable to expect neutrons to be released comparable in energy and abundance with the observed proton tracks.

High energy ($\sim \text{BeV}$) product nucleons (both protons and neutrons) from primary interactions give rise to collision cascades reaching down through the atmosphere, the average energy being degraded by successive collisions which give rise to stars and release more, but lower energy, nucleons. At ~ 100 MeV star production ceases, energy loss by ionization for protons and slowing by elastic and inelastic scattering for neutrons being the predominant processes. The final product of the neutron component which escapes capture comprises slow neutrons (~ 1 eV) in random diffusion in the atmosphere.

* Australian Atomic Energy Commission, University of Melbourne; in conjunction with the University of Ceylon, Colombo, Ceylon.

This picture of the secondary origin of cosmic neutrons has been confirmed by the determination of a primary transition maximum near the top of the atmosphere both for neutron star rates (Camerini *et al.* 1949; Freier, Ney, and Oppenheimer 1949) and the slow neutron intensity (Yuan 1948, 1950).

The present report is concerned with fast and slow neutron fluxes at ground level in the lower half of the atmosphere. Experiments were set up to determine the variation of neutron intensity with altitude at two different latitudes. From the altitude studies the attenuation length of the neutron-producing radiation and related quantities have been derived.

II. LOCATION AND METHOD

The measurements were made in the Garhwal Himalayas during September and October 1951 and in the Australian Alps during January and February 1953. Ilford Nuclear Research emulsions, type C2, 100–200 μ thick were used as detectors. Conditions encountered and the suitability of the Garhwal Himalayas for cosmic ray measurements* have been described (Mather 1952).

The plates were used to make two kinds of measurements :

(A) Determination of relative slow neutron flux by measuring the density of α -tracks from the $^{10}\text{B}(n,\alpha)^7\text{Li}$ reaction, using Ilford C2 plates loaded with 0.023 g cm⁻³ of boron. This reaction generally proceeds to the first excited state of ^7Li leading to an α -track of 1.48 MeV colinear with a recoil of 0.85 MeV giving a track $\sim 7 \mu$ long. A small correction has to be applied for the capture of slow neutrons in nitrogen present in the gelatin, $^{14}\text{N}(n,p)^{14}\text{C}$ (cross section 1.7 barns compared with ~ 740 for the boron reaction), which gives rise to $\sim 6 \mu$ protons not readily distinguished from the α -tracks. This was determined experimentally by exposing unloaded C2 plates under the same conditions.

(B) Direct determination of the high energy flux by observing the production rate of stars per c.c. of emulsion. The majority of evaporation stars at mountain altitudes are due to neutrons of several hundred MeV energy which are themselves products of interactions at higher levels of the collision chain.

Plates were exposed in the Himalayas as follows :

Satpuli..	..	2,100 ft, residual atmosphere	955 g cm ⁻²
Pauri	..	5,300	849
Hanuman Chatti	8,400		757
Badrinath	..	10,400	701
Saraswati Camp	15,800		565
Mana Pass	..	18,400	508

Exposures, including allowance for ascending and descending times, ranged from 1200 to 540 hr. Geomagnetic latitudes were between 19° 30' and 21° 9' N.

* A 17 kW diesel-electric generator is now installed at the town of Badrinath (10,400 ft, geomagnetic lat. 20.5° N.) so that electrical equipment can be operated there.

The same quantities were measured in the Australian Alps, except that some boron-loaded plates were shielded with 0.5 mm of cadmium which absorbs all neutrons $\gtrsim 0.4$ eV energy. The Australian positions were:

Melbourne	sea-level, residual atmosphere	1,030 g cm ⁻²
Mt. Kosciuszko	{ 5,200 ft 7,300	853
area		788

Exposures were ~ 1300 hr, geomagnetic latitude $45^\circ 4'$ S. In all cases exposures were long enough to provide good time averages of neutron intensity. Moreover, as an obvious precaution against local variations of slow neutron flux (Section III), several groups of plates were placed about each site in the hope of averaging out differences.

Very little control could be exercised over temperature in these experiments. At the higher Himalayan sites day-time temperatures in the direct blaze of the Sun were uncomfortably high and would have led to severe fogging of the emulsions. On the other hand night temperatures were below zero. Some obvious precautions were followed. The emulsions were prepared by Ilford Ltd. with extra plasticiser. At the lower altitudes boxes of plates were left in villages under a light covering. At higher altitudes they were given shade by covering with an outer highly reflecting tin cover. In spite of this some of the emulsions showed signs of deterioration, generally in the form of background fog.

To minimize fading of the latent image (which can be troublesome in long exposures) the plates were sealed in containers with silica gel. In addition a number of vacuum-tight aluminium boxes were constructed to contain plates. These were pumped and sealed before leaving Colombo.

III. ORIGIN OF SLOW NEUTRONS AT GROUND LEVEL

First consider the case of neutron measurements in free space (strictly speaking only balloon flights in which a lightly covered detector is suspended well below the balloon). The average neutron energy from evaporation stars is probably ~ 15 MeV. These fast neutrons will be moderated by inelastic and elastic scattering by air atoms, those which survive capture being reduced to ~ 1 eV in ~ 130 collisions in a r.m.s. distance of ~ 150 g cm⁻². The neutron intensity changes by only a factor of ~ 2.5 in this depth so the slow neutron measurements essentially reflect the local neutron production rate in nitrogen and oxygen (Bethe, Korff, and Placzek 1940; Davis 1950).

At the surface of the Earth the situation is considerably modified for two reasons. Firstly, neutrons originating in air within 150 g cm⁻² from the ground may diffuse into the ground where they are moderated and may then escape back to air. Secondly, neutron production occurs in the ground. Neutrons of ~ 15 MeV require perhaps 200 collisions in the ground to slow them and they may then return into air from production depths up to ~ 200 g cm⁻². Both effects will modify (to a degree depending on the atomic composition of the ground surface, and very difficult to estimate) the neutron spectrum and the slow neutron intensity so that the counting rate of a slow neutron detector cannot be related simply to free space measurements.

However, this does not invalidate ground neutron measurements of altitude behaviour provided conditions remain constant. It is well established that neutron production rate measurements in local condensed materials of light elements lead to the same attenuation length values as free space experiments (Simpson 1951; Simpson and Uretz 1953). In the present work the boron-loaded plates were exposed at ground level in topographical settings as comparable as possible. At higher Himalayan sites they were located on extensive outcrops of windswept rock which, during the summer, remained essentially snow-free.

The basic assumption had to be made that the atomic composition of the underlying ground was the same at all sites, which did not appear unreasonable for selected sites in the same mountain range. We believe this assumption has been vindicated by the comparatively small scatter of the results. This is in line with the remarkable uniformity of elemental composition of the lithosphere so far as major constituents are concerned.*

In view of the belief that slow neutrons are final products of star neutrons a direct measurement of star rates (B) can be expected to display approximately the same attenuation with altitude. "Back-streaming" of neutrons of ~ 200 MeV is much less probable and it may reasonably be assumed that most stars will arise from fast neutrons proceeding downwards, making the result less dependent on the proximity of ground.

On the other hand a proportion of emulsion stars are disintegrations of heavy nuclei (silver and bromine) for which anomalous absorption lengths have been reported (e.g. in lead, Simpson and Uretz 1953).

IV. SLOW NEUTRON MEASUREMENTS

We shall consider the altitude study in two parts, the present section dealing with slow neutron production in local ground and air.

Boron-loaded emulsions were scanned at $450\times$ using a C.T.S. microscope with a $45\times$ fluorite objective of N.A. 0.95 which revealed the α -tracks clearly. However, extremely careful scanning was necessary to detect the steeply inclined tracks. The possibility of a small loss is unimportant for relative flux measurements of the kind being reported here but for absolute determinations (for instance, to correlate the α -count with counting rates of boron trifluoride counters) it must be determined. The problem is considered further in Appendix I.

Tracks were counted only if both ends lay inside the emulsion, which resulted in a certain loss from tracks beginning within distance R of either face of the emulsion, where R is the α -range. This factor can be calculated ;

$$F = \frac{\text{tracks counted}}{\text{true number of tracks}} = 1 - \int_0^R \int_0^{\cos^{-1}x/R} \frac{4\pi}{t} \cdot \sin \theta \cdot d\theta \cdot dx$$

$$= 1 - \frac{R}{2t},$$

where t is the emulsion thickness. For a $100\ \mu$ emulsion with $R=7\ \mu$, $F=0.97$.

* Most rocks are dominated by the light elements oxygen, silicon, and aluminium in the following proportions: 48, 28, and 8 per cent.

Figure 1 shows the counting rates for α -tracks from experiments at the two latitudes *v.* residual atmospheric pressure in g cm^{-2} . Assuming that the intensities I may be fitted by an exponential of the form

$$I = I_0 e^{-x/L},$$

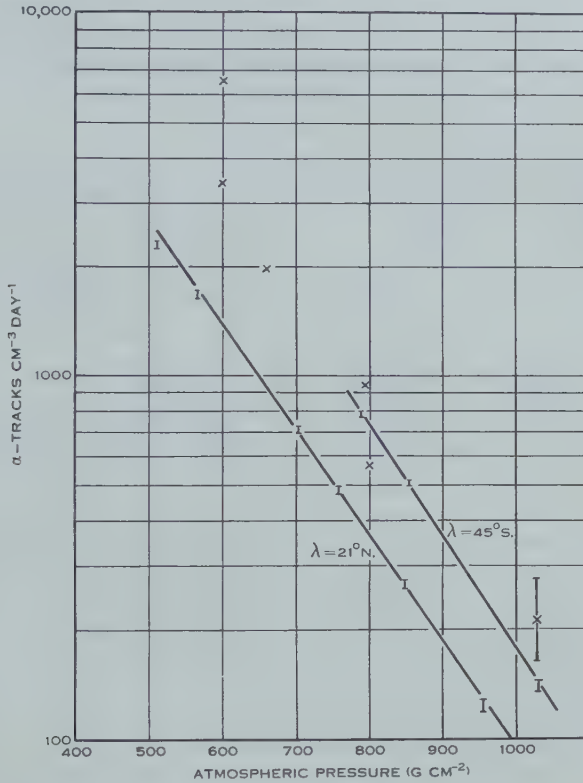


Fig. 1.—Altitude variation of α -tracks in boron-loaded emulsions due to slow neutrons. The upper curve is for $\lambda = 45^\circ \text{S.}$, the lower for $\lambda = 21^\circ \text{N.}$ Points \times are taken from Kaplan and Yagoda (1952) for $\lambda = 50^\circ \text{N.}$

where x is the atmospheric depth and L the characteristic attenuation length, values of L were obtained as follows:

$$\begin{aligned} \text{geomagnetic lat. } 21^\circ \text{ N., } \quad L &= 148 \pm 5 \text{ g cm}^{-2}; \\ 45^\circ \text{ S., } \quad L &= 141 \pm 6 \text{ g cm}^{-2}. \end{aligned}$$

In view of the estimated experimental errors the difference between these is hardly significant although the small increase in L would be consistent with reduced average primary energy at higher latitudes leading to shorter collision chains of the nucleon component in the atmosphere and probably also to a smaller multiplicity of secondary events. Table 1 lists L values measured in the lower atmosphere by several other workers.

There is a tendency in the Himalayan data for the slope to decrease (L increase) for $x \lesssim 600 \text{ g cm}^{-2}$, implying a degradation of the average energy of the neutron-producing radiation as x increases, although the accuracy of the present work is insufficient to prove this. Unfortunately the boron-loaded plates located at the Mana Pass were disturbed several days before the end of the exposure by a band of Tibetan shepherds. They were thrown into snow some distance away and although many of them were still wrapped and proved usable the effect of the change on the local neutron flux is not known. The value for 508 g cm^{-2} is therefore uncertain to this extent.

TABLE 1
SOME CONTEMPORARY L VALUES AT VARIOUS LATITUDES FROM SIMILAR METHODS

Magnetic Latitude (deg.)	L	Method	Reference
0	145	S.n. (C)*	Simpson, Fonger, and Treiman (1953)
21	148	S.n.†	Present paper
21	139	Stars	Present paper
41	144	S.n. (C)	Simpson and Fagot (1953)
45	135	Stars	Bernadini, Cortini, and Manfredini (1950)
45	141	S.n.	Present paper
45	130	Stars	Present paper
46	190	Stars	Lattimore (1949)
46	150	Stars	George and Jason (1949)
50	143	Stars	Yagoda and Kaplan (1949) ; Yagoda, Kaplan, and Conner (1949)
51	160-	S.n.	Kaplan and Yagoda (1952)
52	<148	S.n. (C)	Simpson and Fagot (1953)

* S.n. (C) means detection of slow neutrons produced in local carbon.

† S.n. means detection of slow neutrons.

Nevertheless there is reason to believe that intensity $v.$ altitude curves are not strictly exponentials. The value of L probably decreases slowly with increase in x . Simpson and Fagot (1953) reported a ~ 17 per cent. fall in L from $x \simeq 300$ to 700 g cm^{-2} and concluded that L is a function of depth, decreasing with increase in x from a high altitude value depending on the magnetic latitude λ towards an approximately constant value for all λ , $L \rightarrow 140 \text{ g cm}^{-2}$ for $x \lesssim 600 \text{ g cm}^{-2}$. The present work is not inconsistent with this result. From the latitude measurements of Simpson (1951) and Yuan (1949) at aircraft altitudes ($200\text{--}500 \text{ g cm}^{-2}$) L values of 190 and 170 can be interpolated for latitudes 21 and 45° respectively compared with our measured values of 148 and 141. The evidence from the present work is that L is rather insensitive to λ and tends to a value between 140 and 150 g cm^{-2} .

Boron-loaded plates shielded with cadmium (cut-off $\sim 0.4 \text{ eV}$) showed track densities ~ 0.2 of the unshielded, corresponding to a cadmium ratio of ~ 5 which is considerably greater than observed in free space (~ 2.1 , Yuan

1948; Simpson 1949; and others) indicating, as expected, a greater proportion of slow neutrons due to the presence of the ground. The slow neutron flux (unshielded—cadmium shielded) showed the same exponential absorption with altitude.

The only published data which we can compare directly with our absolute values are those of Kaplan and Yagoda (1952) who used boron and lithium loading in plates on mountains ($\lambda=50^\circ\text{N.}$). Their boron data are shown on Figure 1 (reconverted to α -counts using $\sigma=715$ barns, as adopted by them). Lithium data are not strictly comparable, choice of average cross sections being involved. The conditions of exposure of their plates is not known and precautions may not have been taken against variation in local ground material, which may account for the greater scatter of their points. However, the slope of the intensity-altitude relationship is not inconsistent with ours. Moreover it will be seen that $I_{50} > I_{45} > I_{21}$ in accordance with the well-established latitude variation. No detailed latitude comparison can be effected because of the different longitudes involved.*

Sea-level values over ground have special interest as controls for flux measurements underground and will be referred to in detail. For direct comparison with certain other measurements which have been published we have converted the α -counts to thermal neutron flux ϕ ($=nv$) as follows:

$$\phi = I(\alpha)/\Sigma,$$

where $I(\alpha)$ is the number of α -tracks $\text{cm}^{-2} \text{day}^{-1}$ and Σ the macroscopic cross section $=N_{cc} \cdot \sigma$, N_{cc} being the number of boron atoms per cm^3 in the emulsion (1.28×10^{21}) and σ the cross section for natural boron (18.8 per cent. ^{10}B) is 718 barns for thermal neutrons (2200 m sec^{-1}). If one assumes the neutron spectrum to be thoroughly thermalized and also that the surrounding medium obeys a $1/v$ law, the effective average cross section is $\frac{1}{2}\sqrt{\pi} \cdot \sigma \simeq 600$ barns, giving $\Sigma = 0.77 \text{ cm}^{-1}$. In view of the large cadmium ratio found this may not be too bad an approximation for ground measurements although the shortcomings of the procedure are obvious.

Doing this we obtain as best sea-level flux values ~ 95 (extrapolated) and $\sim 180 \text{ n cm}^{-2} \text{day}^{-1}$ at 21°N. and 45°S. respectively, which may be compared with ~ 270 (corrected to $\sigma=600$ barns) at 50°N. due to Kaplan and Yagoda (1952). These are at least qualitatively consistent with latitude and longitude differences expected for the regions.

Eugster (1954) reported a value of $23 \text{ n cm}^{-2} \text{day}^{-1}$ at Berne, Switzerland, at an altitude of 560 m, equivalent to 966 g cm^{-2} . Correction to sea-level reduces this to ~ 15 which is some 18 times lower than the Kaplan and Yagoda† (1952) value for the same latitude (and the longitude difference should not be

* A neutron minimum probably occurs in the direction in which the equivalent dipole representing the Earth's field is off-centre, viz. $\sim 160^\circ\text{E.}$ long. A correction for longitude difference is expected to be in such a direction as to elevate the Himalayan and Australian data somewhat relative to the North American (Fig. 1).

† An earlier value of ~ 20 published by Yagoda and Kaplan (1949) has been stated by Yagoda (personal communication) to be wrong.

large in this case). Eugster employed boron-loaded emulsions, so his results should be strictly comparable. Local ground fluctuations might account for some of the difference but an order of magnitude effect is highly improbable. Latent image fading during the long exposure he used (116 days) may have been responsible for the low result. This can cause serious loss of tracks unless stringent precautions are taken to minimize it; for instance, exposing the plates in vacuum or a well-desiccated atmosphere. The Swiss experiment was performed at 29 °C and ~ 60 per cent. R.H., the high temperature in particular suggesting severe fading over quite modest periods of time. A not-unrealistic value for the mean life of tracks against fading under these conditions would be ~ 30 days. It would therefore be more appropriate to divide the observed track density not by the exposure period but by the mean life, which would increase the flux estimate by a factor of ~ 4 —at least nearer the mark. It seems probable therefore that sea-level thermal rates for neutrons are in the region $100\text{--}300n\text{ cm}^{-2}\text{ day}^{-1}$ depending on the latitude.

V. STAR RATES

Ilford C2 plates exposed at the above-mentioned sites (Section II) were scanned for stars using a $20\times$ objective. Only stars originating in the emulsion and having three or more prongs were recorded. At least one prong must exceed 50μ to exclude natural α -decay of thorium impurities in the emulsion.

The results are shown in Figure 2 as star population $\text{cm}^{-3}\text{ day}^{-1}$ *v.* altitude for both latitudes. The star rate in the lower half of the atmosphere can be represented satisfactorily by an exponential as in Section IV. Attenuation lengths were calculated to be the following:

$$\lambda = 21^\circ\text{N.}, \quad L = 139 \pm 8\text{ g cm}^{-2};$$

$$\lambda = 45^\circ\text{S.}, \quad L = 126 \pm 15\text{ g cm}^{-2}.$$

There are not enough values from the mountain work at 45° to secure an accurate value for L . The dotted line drawn in Figure 2 attempts to connect the mountain results with a preliminary value ($600\text{ stars cm}^{-3}\text{ day}^{-1}$)* for 43,000 ft (170 g cm^{-2}) obtained by exposing plates for ~ 30 hr in a Canberra jet aircraft flying over the Melbourne area. This is a dubious procedure which we have not yet fully investigated. Star production on mountains may be more influenced by the surrounding material. Moreover, it is clear from Simpson's work referred to in Section IV that the radiation-attenuation is not a simple exponential from sea-level to 200 g cm^{-2} . The mountain data alone ($\lambda = 45^\circ\text{N.}$) would favour $L \simeq 133\text{ g cm}^{-2}$ (full line, Fig. 2).

However, the results generally are consistent with three observations:

- (1) To within experimental uncertainties L values from star rates agree with those from slow neutron fluxes.
- (2) L_{star} tends to be somewhat shorter than L_{neutron} .
- (3) L_{21} is probably larger than L_{45} for stars, as in the case of slow neutrons.

* This value is based on the total time spent over 40,000 ft. The time spent at lower altitudes is not known for this flight so the present value must be regarded as an upper limit.

These facts together confirm the view that stars produced in the atmosphere are the source of the slow neutron component of cosmic radiation. A longer absorption length in each case nearer the equator suggests a higher average energy for the primaries of stars (and hence also the slow neutrons) which is in line with a higher primary cut-off imposed at the equator by the Earth's field. The latitude difference is, however, considerably smaller than observed in the upper half of the atmosphere.

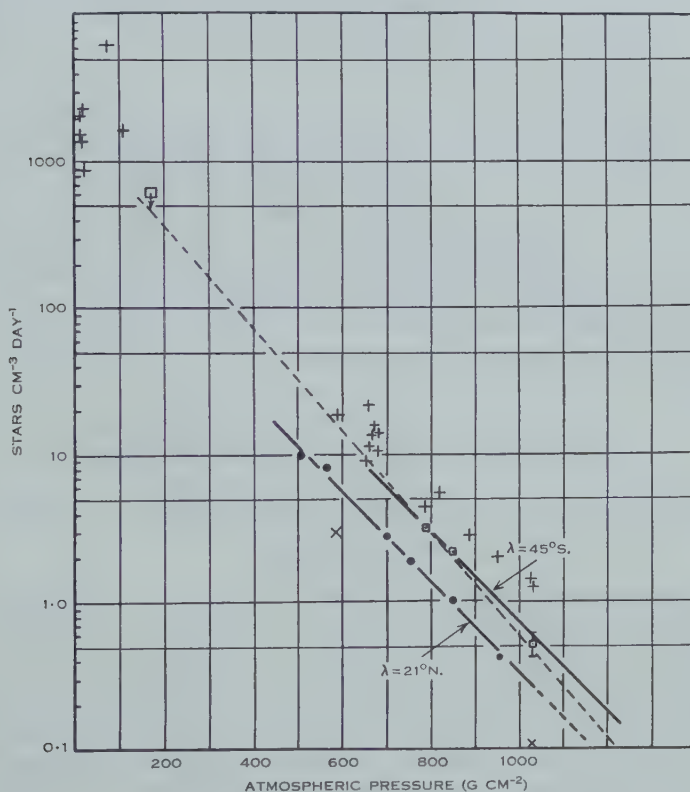


Fig. 2.—Altitude variation of star rates in photographic emulsions. Points \square for $\lambda = 45^\circ\text{S.}$, \bullet for $\lambda = 21^\circ\text{N.}$ Points $+$ are from other workers in the $48\text{--}55^\circ\text{N.}$ latitude range and points \times in the vicinity of $\lambda = 0^\circ$.

On Figure 2 we have also plotted for comparison a selection of star rates observed at other magnetic latitudes. The data between 48 and 54°N. are seen to lie above our 45° line, the difference in absolute star rate being due as much to longitude as latitude difference. This again resembles the regional behaviour of slow neutrons, emphasizing a generic connection.

It is probable that star rates reflect the high energy neutron flux for the following reasons :

(1) The majority of evaporation stars observed had no apparent initiating primary. For C2 emulsions protons of energy $\gtrsim 100$ MeV were not recorded, so

this argument is inconclusive. However, the same thing is observed when G5 emulsions are used, and these are sensitive to minimum ionization tracks.

(2) The average evaporation star energy is ~ 200 MeV excitation energy and a proton of this energy has a range in air of only 20 g cm^{-2} and would therefore be absorbed by ionization loss rather than nuclear collision.

(3) Competing star-producing processes are: $\pi \rightarrow \text{star}$, $\mu \rightarrow \text{star}$, and $\gamma \rightarrow \text{star}$. In the lower half of the atmosphere these are all improbable processes. They have been estimated and discussed by Bernadini, Cortini, and Manfredini (1949, 1950) and Simpson (1951). Not more than 10 per cent. of the stars can arise from these processes.

We now consider the total star rate in a vertical column of atmosphere of unit cross section. As a first approximation we represent the star rate (stars $\text{cm}^{-3} \text{ day}^{-1}$) over the whole depth of the atmosphere at $\lambda = 45^\circ \text{S}$. by

$$I = 2000e^{-x/130}$$

(where $L=130$ has been adopted as probably a close figure for this latitude). We can then calculate the star rate in air from the above rate in emulsion by assuming that the effective cross sections for star production are proportional to the geometric cross sections of the elements involved ($\sigma \propto A^{2/3}$ where A is the atomic weight):

$$I_a \text{ (stars } \text{g}^{-1} \text{ day}^{-1}) = \frac{\bar{A}_e \bar{\sigma}_a}{\rho_e \bar{A}_a \bar{\sigma}_e} \cdot I_e \text{ (stars } \text{cm}^{-3} \text{ day}^{-1}).$$

From the known composition of air and emulsion we find $I_a = 0.41 I_e$, whence

$$I_a = 820e^{-x/130}.$$

Integrating this expression from 0 to 1030 g cm^{-2} gives the total rate of stars occurring in a vertical column of 1 cm^2 . This leads to: $I_a \simeq 1.1 \times 10^5$ stars $\text{cm}^{-2} \text{ day}^{-1}$ or $1.3 \text{ cm}^{-2} \text{ sec}^{-1}$, which is somewhat lower than generally assumed, the difference probably being due to the latitude and longitude of the location. However, for other reasons this value may be somewhat low. The exponential representation ignores the transition effect in star production near the top of the atmosphere. However, taking this figure as an average for the whole atmosphere (the star rate at $\lambda = 45^\circ$ will again be slightly lower than average which probably occurs nearer 50° but the error should be less than ~ 10 per cent.) we can calculate the total star production in the atmosphere of the Earth $= 4\pi R^2 \times 1.3 \simeq 6.5 \times 10^{18}$ stars sec^{-1} or $\sim 1.4 \times 10^{21} \text{ MeV sec}^{-1}$ allotting 220 MeV per average star (R being the Earth's radius).

We can also estimate the intensity of the nucleon radiation responsible for stars at $\lambda = 45^\circ$ on the assumption that the cross section for star production is the geometric one $\sigma = \pi(r_0 A^{1/2})^2$. For an emulsion the weighted average of A is $\bar{A}_e = 30$ and of σ is $\bar{\sigma}_e = 4.5 \times 10^{-25} \text{ cm}^2$ giving 0.65 nucleons $\text{cm}^{-2} \text{ sec}^{-1}$ at the top of the atmosphere (based on $2000 \text{ stars cm}^{-3} \text{ day}^{-1}$) and therefore $0.65e^{-x/130}$ at depth x . This formula will certainly not be true in the transition region but is probably a reasonable approximation below $\sim 150 \text{ g cm}^{-2}$. Moreover, the true cross sections will be somewhat lower than geometric. There is

some evidence to indicate that nuclear transparency is most pronounced for heavy elements such as dominate the emulsion. This makes the present result a lower limit for the true flux. Similar calculations can of course be carried through for the lower latitude, $\lambda=21^\circ\text{N}$. from the data of Figure 2.

The mean geometric cross section $\bar{\sigma}_a=3.5 \times 10^{-25} \text{ cm}^2$ for air corresponds to a mean free path $\sim 70 \text{ g cm}^{-2}$ or rather more if allowance is made for transparency. Since the observed attenuation length is $\sim 130 \text{ g cm}^{-2}$ the primary nucleons of the nucleon component must make on the average rather less than two collisions in air.

VI. STAR MULTIPLICITY

If for any one latitude L values are shorter in the lower atmosphere than the upper and if the transition is gradual (as is presumably the case, corresponding to a degradation of the average energy of the nucleon component) then there should be an accompanying decrease in average star size and energy with atmospheric depth. An analysis is being conducted of star multiplicity at $\lambda=45^\circ\text{S}$. at mountain, aircraft, and balloon altitudes. This is at present incomplete but tentative data seem to bear out the supposition (see Table 2).

TABLE 2
VARIATION OF STAR MULTIPLICITY WITH ALTITUDE
Only evaporation stars are included

Altitude (ft)	Residual Atmosphere (g cm ⁻²)	2-4 Prongs (%)	5-9 Prongs (%)	10 and more Prongs (%)
7,300	788	68.6	27.6	3.8
43,000	170	67	28	5
70,000	46	60.1	33.2	6.7

The mountain data are taken from plates exposed on Mt. Kosciusko, 7300 ft. The values for 43,000 ft are tentative but seem to confirm a relative increase in number of larger stars at the expense of stars in the 2-4 prong class (prong number here includes recoil track if observed). However, the possibility of the difference being due to the presence of ground must be admitted. This might cause an increase in the number of small stars due to back-streaming of secondary neutrons (~ 100 to 150 MeV) from stars produced in the ground. The data for 70,000 ft were kindly supplied by Dr. V. D. Hopper from plates exposed by him in balloon flights. The stars used were all of the evaporation kind having black tracks with some grey tracks but no shower particles.

The 70,000 ft data correspond to the region of the transition maximum and a marked shift from ~ 3 prong ($\sim 150 \text{ MeV}$) to ~ 7 prong ($\sim 300 \text{ MeV}$) and larger stars occurs. While considerably more data are required (especially free-space data in the lower atmosphere) the evidence is generally in favour of a decrease in average star energy with atmospheric depth, consistent with Simpson's conclusions on the trend in L (Simpson and Fagot 1953).

VII. ACKNOWLEDGMENTS

The measurements carried out in the Himalayas were sponsored by the University of Ceylon with the assistance and cooperation of the High Commission for Ceylon in India and many Indian officials to whom the writer acknowledges his gratitude. Professor A. W. Mailvaganam and Dr. V. Appapillai of the University of Ceylon assisted with the planning of the experiment.

The interest of Professor L. H. Martin of the University of Melbourne in the continuation of the work in Australia has been appreciated. Part of the plate scanning was done by Mrs. S. Ovenden.

VIII. REFERENCES

- BERNADINI, G., CORTINI, G., and MANFREDINI, A. (1949).—*Phys. Rev.* **76** : 1792.
 BERNADINI, G., CORTINI, G., and MANFREDINI, A. (1950).—*Phys. Rev.* **79** : 768.
 BETHE, H. A., KOFFE, S. A., and PLACZEK, G. (1940).—*Phys. Rev.* **57** : 573.
 CAMERINI, U., COOR, T., DAVIES, J. H., FOWLER, P. H., LOCK, W. O., MUIRHEAD, H., and TOBIN, N. (1949).—*Phil. Mag.* **40** : 1073.
 DAVIS, W. O. (1950).—*Phys. Rev.* **80** : 150.
 EUGSTER, J. (1954).—*Rev. Sci. Instrum.* **25** : 5.
 FREIER, P., NEY, E. P., and OPPENHEIMER, F. (1949).—*Phys. Rev.* **75** : 1451.
 GEORGE, E. P., and JASON, A. C. (1949).—*Proc. Phys. Soc. Lond. A* **62** : 243.
 KAPLAN, N., and YAGODA, H. (1952).—*Rev. Sci. Instrum.* **23** : 155.
 LATTIMORE, S. (1949).—*Phil. Mag.* **40** : 394.
 MATHER, K. B. (1952).—*Nature* **169** : 616.
 SIMPSON, J. A. (1949).—*Phys. Rev.* **76** : 165.
 SIMPSON, J. A. (1951).—*Phys. Rev.* **83** : 1175.
 SIMPSON, J. A., and FAGOT, W. C. (1953).—*Phys. Rev.* **90** : 1068.
 SIMPSON, J. A., FONGER, W., and TREIMAN, S. B. (1953).—*Phys. Rev.* **90** : 934.
 SIMPSON, J. A., and URETZ, R. B. (1953).—*Phys. Rev.* **90** : 44.
 YAGODA, H., and KAPLAN, N. (1949).—*Phys. Rev.* **76** : 702.
 YAGODA, H., KAPLAN, N., and CONNER, C. H. (1949).—*Phys. Rev.* **76** : 171.
 YUAN, L. C. L. (1948).—*Phys. Rev.* **74** : 504.
 YUAN, L. C. L. (1949).—*Phys. Rev.* **76** : 1267.
 YUAN, L. C. L. (1950).—*Phys. Rev.* **77** : 728.

APPENDIX I

Correction for Scanning Loss

When searching for short tracks there is more chance of failing to observe tracks which are steeply inclined in the emulsion. However, because of the dependence of solid angle on $\cos \delta$, where δ is the dip angle, most tracks appear fairly flat and the chance of loss applies only to the relatively small proportion of very steep tracks. The following measurements were made to determine the actual loss.

In order to obtain a high track density and improve statistics of counting, two boron-loaded C2 200 μ plates were exposed in a bucket of water surrounded by paraffin blocks (both of which served to moderate the neutrons) close to the beryllium target of a 500 keV deuteron electrostatic generator. Emulsion thicknesses were measured just prior to exposure and again after processing and drying to give the shrinkage factor $S=2.69 \pm 2$ per cent.

(i) *Measurement of Track Density.*—The same region of one plate was scanned with a $45\times$ and a $95\times$ objective and a track count made with each using a calibrated eyepiece graticule. The results, based on ~ 2250 tracks at each magnification, were :

$$95\times \quad 6.28 \pm 0.13 \times 10^5 \text{ tracks cm}^{-2},$$

$$45\times \quad 6.34 \pm 0.13 \times 10^5 \text{ tracks cm}^{-2},$$

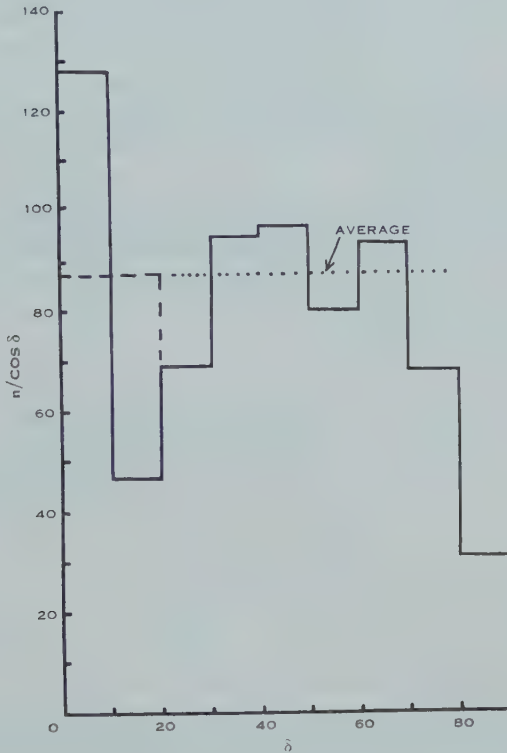


Fig. 3.—Distribution of dip angle δ for α -tracks from $^{10}\text{B}(n, \alpha)^7\text{Li}$ reaction in emulsion.

which agree to within the limits of counting. There is no suggestion of greater loss at $45\times$ than at $95\times$.

(ii) *Distribution of Dip Angles.*—Within a selected area the horizontal and vertical range of every recognizable track was measured using the $45\times$ objective under exactly the same conditions as adopted for the cosmic ray plates. Vertical ranges were corrected for shrinkage ($\times 2.69$) and the dip angles calculated (i.e. δ , the angle of dip formed in the original emulsion). Assuming the distribution of α -tracks to be isotropic in space a plot of $n(\delta)/\cos \delta$ v. δ should be isotropic if no tracks are neglected in scanning. If steep tracks are missed this should be revealed by a fall-off at large δ . The result obtained is shown in Figure 3 based on 500 tracks plotted in 10° blocks. The sharp peak in the first blocks followed by relatively few tracks from 10 – 20° arises from the large depth of field

which is a feature of the $45\times$ fluorite objective. A check on this showed that a single grain could be focused sharply over $\sim 0.7\ \mu$ movement of the fine focus. Now a $7\ \mu$ track having $0.7\ \mu$ vertical dip would have a δ value of 15° . Hence tracks up to $\sim 15^\circ$ appear flat in the field of view and would be recorded with zero dip which would emphasize the 0 – 10° block at the expense of the 10 – 20° block. It is probably a better representation to average over 0 – 20° and this is shown dotted in Figure 3.

The result is a histogram consistent with isotropy from 0 to $\sim 70^\circ$. There is some suggestion of an increase in ordinate from 0 to 50° which has been confirmed by more accurate range-depth measurements with the $95\times$ objective which has a more restricted depth of field. However, this effect, which is probably linked with the choice of S , is not important to the present discussion and will be reported separately.

A decline occurs beyond 70° which has been interpreted as a loss of tracks. The loss factor may be deduced by comparing blocks 70 – 90° with the average (87.1) of the range 0 – 70° (Table 3). The proportion of tracks which should be between δ_1 and δ_2 is $\int_{\delta_1}^{\delta_2} \cos \delta \cdot d\delta$.

TABLE 3
LOSS FACTOR AT DIP ANGLES GREATER THAN 70°

Angle Range (deg)	Tracks Expected (%)	Loss Factor (%)	Absolute Loss of Tracks (%)
70–80	4.5	10.4	0.47
80–90	1.5	64.4	0.97

Hence the total absolute loss of tracks due to scanning by the method used is ~ 1.4 per cent. This correction was applied to the α -density (Fig. 1).

AERIAL SMOOTHING IN RADIO ASTRONOMY

By R. N. BRACEWELL* and J. A. ROBERTS*

[*Manuscript received May 17, 1954*]

Summary

When an aerial is used to survey the distribution of radio brightness over the sky, the observed distribution is smoother than the true distribution; the broader the beam of the aerial, the greater the smoothing. It is shown that the aerial does not register those spatial Fourier components of the true distribution having frequencies beyond a cut-off determined by the aerial aperture. Components of lower frequency are registered but their relative strengths are altered.

Two important consequences follow. (i) There are *invisible distributions* which produce no response when scanned by the aerial. Consequently there is not a unique solution to the problem of correcting for aerial smoothing. The established method of correcting by successive smoothing, leads to the *principal solution*, in which Fourier components accepted by the aerial have been restored to their full values, but the components rejected by the aerial are still not represented. (ii) In conducting a survey it is sufficient to observe at discrete intervals. The measuring points must be closer together than half the period of the Fourier component at cut-off. For an aperture of width w , this *peculiar interval* is equal to $\frac{1}{2}\lambda/w$ (radians).

I. INTRODUCTION

It is the objective of much radio-astronomical work to determine the intensity of the radio waves arriving at the Earth from different directions. For any one frequency and polarization the energy flux of this radiation field is conveniently specified by a distribution of brightness temperature over the celestial sphere. Such a description is appropriate since, in the régime where the Rayleigh-Jeans law holds, the brightness of a black body is directly proportional to the temperature.

One may be interested in the distribution of brightness temperature over the whole of the celestial sphere, a small area the size of the Sun, or, in the case of "radio stars", over still smaller areas. In all cases, the aerials used tend to blur the detail of the distribution. The energy received when an aerial is pointed in a given direction is determined not by the brightness temperature in that direction alone, but by a weighted mean of the temperatures in all the directions contained within the aerial beam. If the aerial pattern is fine enough compared with the scale of the temperature distribution studied, then the observed distribution will approximate closely to the true distribution. Often this is not the case, and then the question arises whether, from an accurate knowledge of the aerial response, it is not possible to reconstruct the original distribution.

* Division of Radiophysics, C.S.I.R.O., University Grounds, Sydney.

It is known that certain types of smoothing process are reversible. For example, it is possible to recover daily sunspot numbers in full detail from daily values of the much smoother 30-day running means. The means must, however, be given without approximation. Aerial smoothing is so closely allied to smoothing by running means that one might surmise that the limit to detail obtainable from a survey with a given aerial could be improved indefinitely by increasing the accuracy of the observations and the closeness of their spacing.

In a number of investigations (Hey, Parsons, and Phillips 1948; Bolton and Westfold 1950) radio astronomers have attempted to reconstruct the original distribution which gave rise to their observations. A proposed restoration may be tested by smoothing it artificially with the known aerial directional diagram, and comparing the result with the observations. Good agreement is usually obtainable, and it has been thought that a restoration, so tested, must be a good approximation to the true distribution. This assumption now proves to be wrong, for it will be shown that there are always an infinite number of different restorations, all of which satisfy the test exactly.

A clear insight into the nature of the blurring produced by the aerial is obtained by studying the effect of the aerial on the spatial Fourier components of the temperature distribution. The true temperature distribution is supposed analysed into Fourier components, each being a sinusoidal temperature distribution of a certain strength and spatial frequency. The aerial affects these components by *eliminating* a semi-infinite band of high frequencies and altering the relative strengths of the lower frequency components. Thus, in the observed distribution the finest detail is irretrievably lost, while the less fine detail may be substantially modified.

The rejection of a semi-infinite band of high frequencies by the aerial is responsible for the existence of *invisible distributions*. These are distributions containing only frequencies which would be rejected by the aerial. They produce no response when scanned by the aerial. There is a great variety of such invisible distributions and consequently a great variety of distributions, each of which when smoothed by the aerial directional diagram reproduces the observed distribution. These possible solutions differ from each other only by the addition of invisible distributions.

To emphasize the importance of this effect two examples from the literature are shown in Figure 1. The radial distribution of brightness across the Sun at 21 cm obtained by Christiansen and Warburton (1953) is shown in Figure 1 (*a*). The authors state that, on the basis of their observations, they could not distinguish between the two very different distributions represented. In the second example (Fig. 1 (*b*)) the full line gives the distribution of 100 Mc/s cosmic noise along the galactic meridian passing through the galactic centre, as deduced by Bolton and Westfold (1950) from their measurements. In the same figure the broken line shows a (renormalized) distribution which according to Brown and Hazard (1953) gives the same result, when scanned with the aerial, as does the full curve.

This indeterminacy in the restoration of the observations made with aerials of finite resolving power seriously affects several important current investiga-

tions. One of these is the study of solar limb brightening, where there is a need for observations to compare with existing theories (e.g. Smerd 1950). As is suggested by Figure 1 (a), the peak brightness at the limb is not an appropriate quantity for comparison of theory and observation. Some sort of integral is more suitable. In another important investigation (Piddington 1951) the brightness of the galactic centre is taken from surveys on different frequencies over a 30 : 1 range, in order to ascertain a spectrum. The correction for aerial smoothing may here be of vital importance. The same serious problem would arise in determining the spectra of discrete sources from attempted measurements of the central brightness.

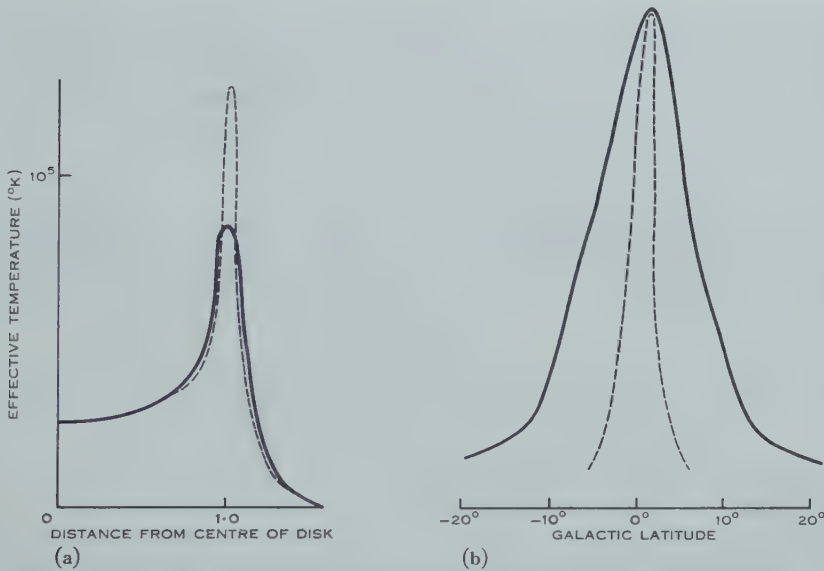


Fig 1 (a).—Solar distributions of brightness temperature (Christiansen and Warburton 1953) which could not be distinguished after subjection to aerial smoothing.

Fig. 1 (b).—A pair of similarly indistinguishable galactic distributions (Brown and Hazard 1953).

Given only the observations made with a finite aerial it is impossible to decide which of the feasible solutions is the true distribution. However, of those distributions which reproduce the observations when scanned by the aerial, the one which contains no invisible part has a certain uniqueness and is termed the *principal solution*. It is this distribution which is found when the observations are corrected by the process of successive substitutions (Section VIII). When the true distribution is well resolved by the aerial, the principal solution is a good approximation. However, if the true distribution is not well resolved, the principal solution may be a poor approximation, and even physically impossible.

To improve on the principal solution one must take into account further information about the distribution which may be available in particular cases. How best to do this in each case is an outstanding problem.

In Section VI we discuss a by-product of this investigation, namely, an important theorem which arises from the rejection of high frequencies by the aerial. There exists a critical angle, peculiar to each aerial, and such that no increase in information is obtained by making observations at finer intervals. In fact, from observations spaced at the peculiar interval, it is possible to infer the intermediate observations. This theorem is expected to be valuable both in the making and in the reduction of observations.

II. BASIC FORMULA

For the present purpose of studying the effects of aerial smoothing and the possibilities of restoration, we shall discuss the simplest case, namely, when the brightness temperature $T(\varphi)$ depends on a single angular coordinate φ only. This case includes the essential physics of aerial smoothing and in fact covers many of the practical applications in radio astronomy.

It is convenient to specify the directional characteristics of the aerial by the power response \mathbf{A} to a point source. Thus when the aerial beam is directed towards $\varphi = \varphi_0$ the response of the aerial to a point source at $\varphi = \beta$ is by definition proportional to $\mathbf{A}(\varphi_0 - \beta)$. \mathbf{A} is normalized so that

$$\int \mathbf{A}(\varphi) d\varphi = 1.$$

It follows that the measured temperature distribution, $T_a(\varphi_0)$, corresponding to a point source of temperature T situated at $\varphi = \beta$ is given by

$$T_a(\varphi_0) = \mathbf{A}(\varphi_0 - \beta) T.$$

The response to a point source, $\mathbf{A}(\varphi)$, is the power directional diagram, $A(\varphi)$, taken with φ in the opposite sense, i.e. $A(\varphi) = \mathbf{A}(-\varphi)$. For the development of the theory \mathbf{A} proves to be more convenient than A , for reasons explained below. In many practical cases the directional diagram is symmetrical, so that the distinction need not be drawn.

When now the true temperature distribution $T(\varphi)$ is observed with the aerial pointed in the direction φ_0 , the measured temperature $T_a(\varphi_0)$ will be a weighted mean of T given by

$$T_a(\varphi_0) = \int \mathbf{A}(\varphi_0 - \beta) T(\beta) d\beta.$$

When the aerial is reasonably directive $\mathbf{A}(\varphi)$ becomes negligible beyond a small range of φ and the limits of integration may be extended to $\pm \infty$. For convenience, we also drop the subscript "0", and thus obtain the equation

$$T_a(\varphi) = \int_{-\infty}^{\infty} \mathbf{A}(\varphi - \beta) T(\beta) d\beta. \quad \dots\dots\dots (1)$$

This equation expresses in its simplest form the phenomenon of aerial smoothing and forms the basis of the present communication. In other circumstances, the expression on the right-hand side of (1) would be known as the

convolution, composition product, Faltung, fold, etc. of $\mathbf{A}(\varphi)$ and $T(\varphi)$ or as the (unnormalized) cross-correlation function of $\mathbf{A}(-\varphi)$ and $T(\varphi)$. To represent the convolution of two functions we shall use the abbreviated notation

$$\mathbf{A} * T \equiv \int_{-\infty}^{\infty} \mathbf{A}(\varphi - \beta) T(\beta) d\beta \equiv \int_{-\infty}^{\infty} T(\varphi - \beta) \mathbf{A}(\beta) d\beta,$$

and in this notation

$$T_a = \mathbf{A} * T.$$

We may also write

$$T_a = A \star T,$$

where†

$$A \star T \equiv \int_{-\infty}^{\infty} A(\beta - \varphi) T(\beta) d\beta \equiv \int_{-\infty}^{\infty} T(\beta + \varphi) A(\beta) d\beta.$$

In statistics $A \star T$ would be called the (unnormalized) cross-correlation function of A and T .

III. FORMAL SOLUTION BY FOURIER TRANSFORMS

Equations of the same type as (1) arise frequently in physics (see e.g. Trumpler and Weaver (1953) for references). Typical instances are the instrumental blurring of (i) optical spectra (van Cittert 1931; Burger and van Cittert 1932; van de Hulst 1941, 1946), (ii) X-ray spectra (Stokes 1948; Paterson 1950; Waser and Schomaker 1953), and (iii) solar limb darkening curves (Fellgett and Schmeidler 1952). A good survey of various methods of dealing with two-dimensional examples arising in astronomy is given by Burr (1955).

In the present application we make use of the special properties arising from the restriction of $\mathbf{A}(\varphi)$ to be the response of an aerial to a point source. For this class of problem the approach from Fourier theory is most enlightening. In this procedure the temperature distributions and the aerial pattern are regarded as built up of a spectrum of components harmonic in the angular variable φ . Thus $T_a(\varphi)$ is specified by its spectrum $\bar{T}_a(s)$ which is the (complex) Fourier transform of $T_a(\varphi)$ and gives the amplitude and phase of the harmonic component with s wave crests per unit of φ .

When functions are related as in equation (1) the corresponding relation between their Fourier transforms is very simple. This relation is given by the convolution theorem (Sneddon 1951) which can be stated in the form:

† Because the operation of convolution (denoted by $*$) is commutative ($f * g = g * f$), associative ($f * [g * h] = [f * g] * h$), and distributive ($[f + g] * h = f * h + g * h$) (Doetsch 1937), it may be treated algebraically like ordinary multiplication and thus leads to simple mathematics (Section VIII). On the other hand the operation of smoothing T with A (written $A \star T$) is perhaps more direct for some purposes than forming the convolution $\mathbf{A} * T$ of \mathbf{A} with T . In either case it is A which is plotted or tabulated when the calculation is performed. And, of course, A is the customary quantity in aerial physics, not \mathbf{A} . However, the smoothing process, being non-commutative and non-associative, proves to be not as convenient as convolution in the type of analysis occurring in the present paper.

If $\bar{T}_a(s)$ is the Fourier transform of $T_a(\varphi)$ defined by

$$\bar{T}_a(s) = \int_{-\infty}^{\infty} T_a(\varphi) e^{-i2\pi s \varphi} d\varphi,$$

and, if $\bar{T}(s)$, $\bar{\mathbf{A}}(s)$ are similarly the transforms of $T(\varphi)$ and $\mathbf{A}(\varphi)$, then, when

$$T_a = \mathbf{A} * T,$$

the transforms are related by

$$\bar{T}_a(s) = \bar{\mathbf{A}}(s) \bar{T}(s). \quad \dots\dots\dots (2)$$

To state this theorem in words: if $T_a(\varphi)$ is obtained from $T(\varphi)$ by convolution with $\mathbf{A}(\varphi)$, then the Fourier transform of $T_a(\varphi)$ is the algebraic product of the transforms of $T(\varphi)$ and $\mathbf{A}(\varphi)$.

Equation (2) emphasizes the importance of zeros in the Fourier transform of the aerial diagram. Since \bar{T}_a is obtained from \bar{T} by multiplication with $\bar{\mathbf{A}}$, it follows that for those frequencies s_k for which

$$\bar{\mathbf{A}}(s_k) = 0,$$

the value of \bar{T} is lost in the smoothing process. For example, if $\bar{T}(s)$ contained a "spectral line" at $s = s_k$, $\bar{T}_a(s)$ would contain no trace of it.

For values of s such that $\bar{\mathbf{A}}(s) \neq 0$, the Fourier transform of the true distribution is given by

$$\bar{T}(s) = \frac{\bar{T}_a(s)}{\bar{\mathbf{A}}(s)}.$$

But when $s = s_k$, making both $\bar{\mathbf{A}}$ and \bar{T}_a zero, equation (2) shows that \bar{T} is indeterminate. Thus when $s = s_k$ equation (2) is satisfied not only by $\bar{T}(s)$ but also by

$$\bar{T}(s) + \sum_k a_k \delta(s - s_k),$$

where the a_k are arbitrary and $\delta(s - s_k)$ is unit impulse at $s = s_k$. It follows that $T(\varphi)$ is not the only solution of the integral equation; it is also satisfied by

$$T(\varphi) + \sum_k a_k e^{i2\pi s_k \varphi}.$$

(If $\bar{\mathbf{A}}(s)$ is zero, not at discrete points, but over a continuous range, the summation is replaced by an integration.) The additive functions $\sum a_k \exp(i2\pi s_k \varphi)$, which we term *invisible distributions* for the aerial, are obviously solutions of the integral equation

$$\int_{-\infty}^{\infty} \mathbf{A}(\varphi - \beta) T(\beta) d\beta = 0.$$

They are of such a nature that it is impossible to detect them with the aerial in question whatever their magnitude (see Section VI).

It is thus clear that the zeros of the transform of $\mathbf{A}(\varphi)$ play a vital role in the theory. For this reason it is now necessary to investigate the Fourier transforms of aerial patterns.

IV. AERIAL THEORY

In considering the nature of the Fourier transforms of aerial patterns in this section, we first illustrate their characteristics by discussion of a common type of aerial, and then establish a general theorem. For this purpose we regard the aerial as used for transmission or reception, whichever gives the simplest description, since the directional diagrams are identical in the two cases.

Thus for illustration consider an aerial consisting of a finite one-dimensional aperture of width w , across which is maintained a field constant in amplitude and phase. For this aerial the $\mathbf{A}(\varphi)$ for small φ is approximately

$$\mathbf{A}(\varphi) = \frac{\lambda}{w} \left\{ \frac{\sin(\pi\varphi w/\lambda)}{\pi\varphi} \right\}^2,$$

the numerical factor being chosen so that $\int_{-\infty}^{\infty} \mathbf{A}(\varphi) d\varphi = 1$. $\mathbf{A}(\varphi)$ is shown in

Figure 2 (a). The beam width of the main lobe between zeros is $2\lambda/w$, the width to half power is $0.89\lambda/w$. Figure 2 (b) shows $\bar{\mathbf{A}}(s)$, the Fourier transform of

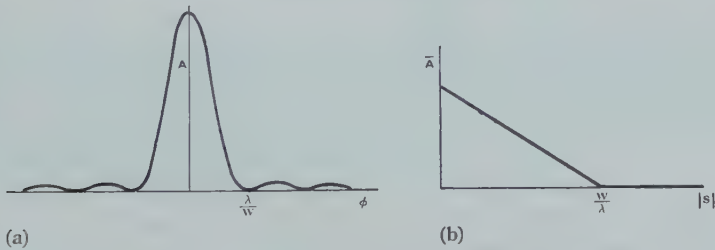


Fig. 2.—The response to a point source, $\mathbf{A}(\varphi)$, and its Fourier transform $\bar{\mathbf{A}}(s)$ for a uniformly illuminated aperture.

$\mathbf{A}(\varphi)$. Since $\bar{\mathbf{A}}(s)$ is real and symmetrical in this case, it may conveniently be plotted against $|s|$. (In subsequent figures illustrating transforms, we have, where convenient, taken them real and symmetrical.) The particular feature to notice is that $\bar{\mathbf{A}}(s)$ is zero for all values of s greater than the limiting frequency $s_c = w/\lambda$. We shall now show that this feature is common to all ordinary aerials consisting of finite plane apertures.

It is known (Booker and Clemmow 1950) that a one-dimensional aperture distribution of electric field $E(x/\lambda)$ and the angular spectrum $P(\sin \varphi)$ to which it gives rise are reciprocal Fourier transforms with respect to $\sin \varphi$ and x/λ , that is,

$$P(\sin \varphi) = \int_{-\infty}^{\infty} E\left(\frac{x}{\lambda}\right) e^{i2\pi(x/\lambda) \sin \varphi} d\left(\frac{x}{\lambda}\right),$$

and

$$E\left(\frac{x}{\lambda}\right) = \int_{-\infty}^{\infty} P(\sin \varphi) e^{-i2\pi(x/\lambda) \sin \varphi} d(\sin \varphi).$$

The angular spectrum $P(\sin \varphi)$ is proportional to the field directional diagram for real values of φ . When $|\sin \varphi| > 1$, that is, for imaginary values

of φ , $P(\sin \varphi)$ represents the evanescent field of the aerial, which is normally small for very directive aerials.* If we assume it to be negligible the power directional diagram $A(\varphi)$ is given by

$$\begin{aligned} A(\varphi) &= \text{const.} \cdot |P(\sin \varphi)|^2 \\ &= \text{const.} \cdot P \dagger P, \quad \dots \dots \dots (3) \end{aligned}$$

where $P \dagger$ is the complex conjugate of $P \equiv P(\sin \varphi)$. The directional diagram $A(\varphi) = \mathbf{A}(-\varphi)$ is used here instead of $\mathbf{A}(\varphi)$, the response to a point source, to obtain agreement with the usual practice in aerial theory.

As a further consequence of the assumption of directive aerials, it is permissible to replace $\sin \varphi$ by φ in the equations above, since $P(\sin \varphi)$ is negligible where this approximation is invalid. Then applying the convolution theorem of Section III to equation (3) we have

$$\bar{A}(s) = \text{const.} \cdot \bar{P} \dagger * P,$$

and with a little further reduction

$$\bar{\mathbf{A}}(s) = \text{const.} \cdot E \star E \dagger.$$

Stating this important result in words, the Fourier transform of the aerial response pattern $\mathbf{A}(\varphi)$ is proportional to the (complex) auto-correlation function of the aperture distribution (see also Booker, Ratcliffe, and Shinn 1950).

For an aperture of width w , it follows that $\bar{\mathbf{A}}(s)$ is zero for frequencies greater than w/λ . This result is in obvious agreement with the discussion of the uniformly fed aperture above, but it extends this result to more general aerials.

We have not covered all aerials but we know from experience that the directivity of a broadside array cannot be exceeded very much by rearranging its elements within the same overall dimensions. In any particular case the existence of the cut-off could be verified and the value of s_c determined by taking the Fourier transform of the aerial diagram.

The highest frequency which can be present in $\mathbf{A}(\varphi)$, the cut-off frequency s_c , is given by

$$s_c = \frac{w}{\lambda}.$$

The inverse of this frequency

$$\varphi_c = \frac{1}{s_c}$$

may be termed the cut-off period. For a uniformly illuminated aperture it is half the beam width between zeros or 1.12 times the beam width to half power.

V. AERIAL SMOOTHING

It has been shown in Section III, equation (2), that the effect of aerial smoothing may be described in terms of spectra by the statement that the spectrum of the observed temperature distribution T_a is obtained from the

* Very directive aerials with large evanescent fields have been mooted (see, for example, Woodward and Lawson 1948) but are of more theoretical than practical interest.

spectrum of the true distribution T by multiplication with the spectrum of the response of the aerial to a point source. But, according to the result of Section IV, the spectrum of any aerial response vanishes for all frequencies s beyond a limiting value s_c . The high frequency components of T are thus entirely rejected by the radiometer.

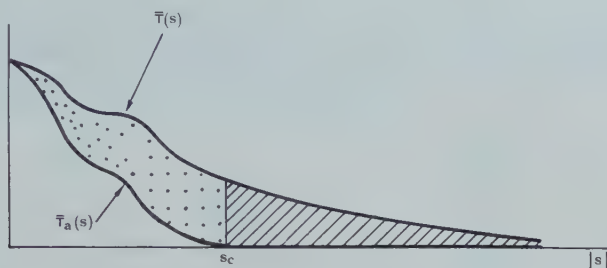


Fig. 3.—Showing how the spectrum of a smoothed distribution, $\bar{T}_a(s)$, is related to the spectrum of the original distribution $\bar{T}(s)$.

This is illustrated in Figure 3. $\bar{T}(s)$ is the spectrum of some $T(\varphi)$ which includes high frequencies. In $\bar{T}_a(s)$, the spectrum of the observed distribution, the whole of $\bar{T}(s)$ for frequencies greater than s_c (cross-hatched area) has been lost, and furthermore the lower frequency components have been reduced by the loss of the stippled area.

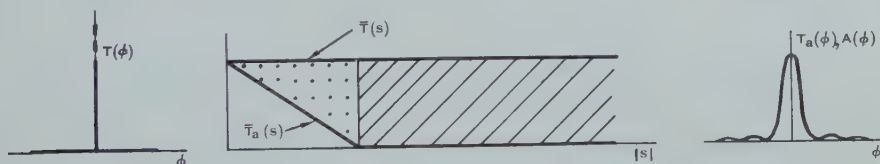


Fig. 4.—The effect of scanning a point source illustrated in terms of transforms.

Figure 4 gives the corresponding diagram for a point source scanned by an aerial consisting of a uniformly illuminated aperture for which

$$\mathbf{A}(\varphi) = \frac{1}{\pi} \left(\frac{\sin \varphi}{\varphi} \right)^2.$$

Since $T(\varphi)$ is an impulse function, its spectrum $\bar{T}(s)$ is flat (see Fig. 4). $\bar{T}_a(s)$, which contains none of the higher frequencies, and in which the lower frequencies have been reduced, is triangular as shown. $T_a(\varphi)$ has the same shape as $\mathbf{A}(\varphi)$, as is obvious.

Further illustrations of the spectral aspect of smoothing are given in Figure 5, which shows two cases of $T(\varphi)$, a rectangular distribution and a Gaussian error curve. Both true distributions are sufficiently broad, compared with the aerial beam, to be fairly well resolved.

In all cases $T_a(\phi)$ is a distribution which contains no frequencies greater than a certain limit. This feature has important repercussions on the possibilities of restoration on the one hand and on the other hand has a practical bearing on the conduct of surveys. In the following section we discuss the properties of such "band-limited" functions.

VI. A THEOREM CONCERNING OBSERVED DISTRIBUTIONS

Observed distribution functions $T_a(\phi)$, by virtue of containing no frequencies greater than a limiting value s_c , have a property which is of great practical

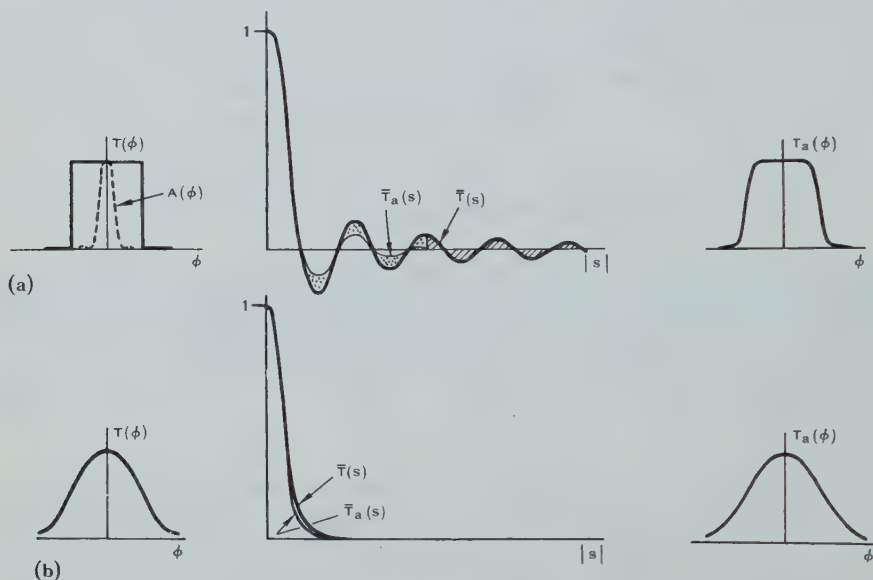


Fig. 5.—The effect of scanning (a) a rectangular and (b) a Gaussian source with the aerial pattern $A(\phi)$.

significance. *An observed distribution is completely determined by measurements spaced at equal discrete intervals which are at least as narrow as $\frac{1}{2}\phi_c = \frac{1}{2}s_c^{-1}$.* The interval $\frac{1}{2}\phi_c$ is peculiar to the aerial used for the observations. For example, an aerial consisting of an aperture of overall width w has a peculiar interval of $\frac{1}{2}\lambda w^{-1}$, which in the case of a uniformly fed aperture is numerically equal to 0.56 times the beam width to half power. In this latter case also it is worth mentioning that the peculiar interval is just half the Rayleigh limit of resolution.

The peculiar interval theorem is analogous to one now well known in communication theory (see e.g. Shannon 1949). The proof given below is novel.

For the discussion of functions sampled at equal intervals of ϕ , it is convenient to define the improper function $\text{III}(\phi)$ which consists of an infinite sequence of unit impulse functions at unit separation, that is,

$$\text{III}(\phi) = \sum_{n=-\infty}^{\infty} \delta(\phi - n).$$

This useful function is the same as the "row of impulses" of Heaviside (1922), who expressed it as the limit of a Fourier series. We find the symbol III convenient and pronounce it *shah* after the Cyrillic character. The function III has the simple and important property that its Fourier transform is also III . This is easily found by taking the limit of the Fourier series for a rectangular waveform. The transform of $\text{III}(\varphi/b)$ is $b\text{III}(bs)$. A further notation which is convenient when dealing with cut-off functions is $\Pi(s)$ for the even rectangle function of unit height and base; $\Pi(s)$ is unity for $-\frac{1}{2} < s < \frac{1}{2}$, zero for $|s| > \frac{1}{2}$, and $\Pi(\pm\frac{1}{2}) = \frac{1}{2}$. The function $\Pi(s/b)$ is a rectangle of unit height and base b and its Fourier transform is $(\pi\varphi)^{-1} \sin \pi b\varphi$.

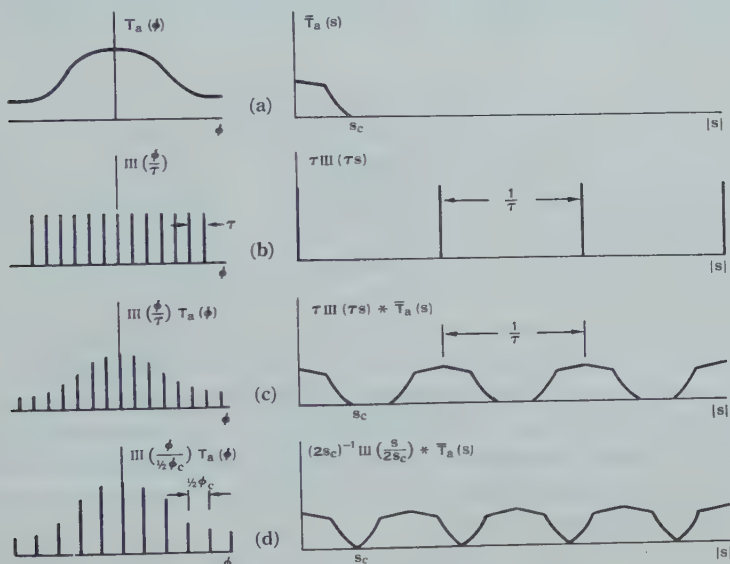


Fig. 6.—Several functions and their spectra, (a) an observed distribution, (b) a special sampling function, (c) sampling the distribution at discrete intervals is equivalent to multiplication by the sampling function, (d) sampling at the peculiar interval.

Samples taken at intervals τ from a function $T_a(\varphi)$ contain the same information as $\text{III}(\varphi/\tau)T_a(\varphi)$. This function, which is a sequence of impulses at intervals τ and of strength proportional to $T_a(\varphi)$, is illustrated on the left in Figure 6 (c), impulse functions being shown as finite spikes of height proportional to their integral.

The Fourier transform of $\text{III}(\varphi/\tau)T_a(\varphi)$ is $\tau\text{III}(\tau s)*\bar{T}_a(s)$, that is, a function formed by repeating $\bar{T}_a(s)$ at intervals τ^{-1} . This follows from the convolution theorem, or it may be verified directly by calculating the Fourier coefficients of the periodic function $\text{III}(\tau s)*\bar{T}_a(s)$. If T_a is sampled at sufficiently small intervals (τ sufficiently small) the spectrum of $\text{III}(\varphi/\tau)T_a(\varphi)$ will consist of well-separated parts as shown on the right of Figure 6 (c). It is evident that, as long as τ^{-1} is greater than $2s_c$, that is, while the sampling interval τ is less than $\frac{1}{2}\varphi_c$, the separate parts will not overlap. Hence, when $\tau < \frac{1}{2}\varphi_c$, the exact spectrum

of T_a is deducible from the spectrum of $\text{III}(\varphi/\tau)T_a(\varphi)$, and therefore, finally, T_a itself is deducible from the values of T_a at intervals τ .

To recover T_a from $\text{III}(\varphi/\tau)T_a(\varphi)$, it is merely necessary to remove all the Fourier components with frequencies beyond s_c , that is, to multiply the spectrum of $\text{III}(\varphi/\tau)T_a(\varphi)$ by $\Pi(s/2s_c)$. By the convolution theorem this is equivalent to smoothing $\text{III}(\varphi/\tau)T_a(\varphi)$ with $(\pi\varphi)^{-1} \sin 2\pi s_c \varphi$. Since $\text{III}(\varphi/\tau)T_a(\varphi)$ is a sequence of impulse functions, the smoothing integral reduces, without approximation, to a summation.

In practice, interpolation midway between values of T_a given at the peculiar interval $\frac{1}{2}\varphi_c$ is the most frequently needed process. To do this, first prepare a table in which T_a appears in a column at unit intervals of $\Phi \equiv \varphi/\frac{1}{2}\varphi_c \equiv 2s_c\varphi$. Then tabulate $(\pi\Phi)^{-1} \sin \pi\Phi$ for $\pm\frac{1}{2}$, $\pm 1\frac{1}{2}$, $\pm 2\frac{1}{2}$, etc. in a strip thus:

1273
<u>2122</u>
6366
—→
6366
<u>2122</u>
1273

the spacing being the same as in the table of T_a . The decimal points are omitted and negative values are indicated by bars. Table 1 gives further values. Place the strip alongside the T_a column, and the sum of the products of the adjacent numbers then gives the interpolated value of T_a opposite the arrow. To establish this procedure, put $\Phi = \text{integer} + \frac{1}{2}$ and $\tau = \frac{1}{2}s_c^{-1}$ in the general equation for recovery of T_a . Thus

$$\begin{aligned}
 T_a &= \frac{\sin 2\pi s_c \varphi}{\pi \varphi} * \text{III}\left(\frac{\varphi}{\tau}\right) T_a(\varphi) \\
 &= \frac{\sin 2\pi s_c \varphi}{2\pi s_c \varphi} * 2s_c \text{III}(2s_c \varphi) T_a(\varphi) \\
 &= \sum_{\pm\Phi = \frac{1}{2}, 1\frac{1}{2}, \dots} \frac{\sin \pi\Phi}{\pi\Phi} T_a\left(\frac{1}{2}\varphi_c \Phi - \varphi\right).
 \end{aligned}$$

This property of the observed distributions, that they are sufficiently specified by spot values at intervals, is a direct consequence of the nature of aerial diagrams, and is obviously of great importance. For the observational programme it implies that (apart from the reduction of errors) nothing is gained by taking readings at intervals closer than $\frac{1}{2}\varphi_c = \frac{1}{2}s_c^{-1} = \frac{1}{2}\lambda w^{-1}$.

For computational work the property is equally important as it permits a band-limited function to be represented exactly by a set of discrete values. To ensure that data used in numerical work are in fact free from high frequencies, the data may be first "filtered" by smoothing with

$$\frac{\sin \pi\Phi}{\pi\Phi}.$$

TABLE 1
INTERPOLATING FUNCTION

$ \Phi $	$\frac{\sin \pi\Phi}{\pi\Phi}$	$ \Phi $	$\frac{\sin \pi\Phi}{\pi\Phi}$	$ \Phi $	$\frac{\sin \pi\Phi}{\pi\Phi}$	$ \Phi $	$\frac{\sin \pi\Phi}{\pi\Phi}$	$ \Phi $	$\frac{\sin \pi\Phi}{\pi\Phi}$
$\frac{1}{2}$	0.6366	$7\frac{1}{2}$	-0.0424	$14\frac{1}{2}$	0.0220	$21\frac{1}{2}$	-0.0148	$28\frac{1}{2}$	0.0112
$1\frac{1}{2}$	-0.2122	$8\frac{1}{2}$	0.0374	$15\frac{1}{2}$	-0.0205	$22\frac{1}{2}$	0.0141	$29\frac{1}{2}$	-0.0108
$2\frac{1}{2}$	0.1273	$9\frac{1}{2}$	-0.0335	$16\frac{1}{2}$	-0.0193	$23\frac{1}{2}$	-0.0135	$30\frac{1}{2}$	0.0104
$3\frac{1}{2}$	-0.0909	$10\frac{1}{2}$	0.0303	$17\frac{1}{2}$	-0.0182	$24\frac{1}{2}$	0.0130	$31\frac{1}{2}$	-0.0101
$4\frac{1}{2}$	0.0707	$11\frac{1}{2}$	-0.0277	$18\frac{1}{2}$	0.0172	$25\frac{1}{2}$	-0.0125	$32\frac{1}{2}$	0.0098
$5\frac{1}{2}$	-0.0579	$12\frac{1}{2}$	0.0255	$19\frac{1}{2}$	-0.0163	$26\frac{1}{2}$	0.0120	$33\frac{1}{2}$	-0.0095
$6\frac{1}{2}$	0.0490	$13\frac{1}{2}$	-0.0236	$20\frac{1}{2}$	0.0155	$27\frac{1}{2}$	-0.0116	$34\frac{1}{2}$	0.0092

VII. POSSIBILITIES OF RESTORATION

Having studied the nature of $\bar{\mathbf{A}}(s)$ for practical aeriels we are now in a position to consider the solution of the integral equation (1). It follows immediately from the general discussions of Section III that there is no unique solution. From any known solution an infinite set of further solutions can be found by adding one of the invisible distributions. These are of the form $ff(s) \exp(i2\pi\phi s)ds$, where $f(s)$ is arbitrary and the integration extends over any values of s for which $\bar{\mathbf{A}}(s)$ is zero. These distributions produce no output when scanned by the aerial.

An infinite variety of such invisible distributions can be constructed. Those illustrated in Figure 7 are as follows: a harmonic distribution with spatial frequency beyond the cut-off frequency s_c ; a "wave packet" containing only spectral components beyond cut-off; an impulse function from which the low frequencies have been removed; and a step distribution from which low frequencies have been removed.

In Figure 8 we have several solutions to a particular problem obtained by adding some of these distributions to an already known solution.

Since the invisible distributions contain no zero frequency component, they must all contain negative as well as positive values. In fact, to ensure that the aerial receives nothing, they must oscillate from positive to negative within the space of about one beam width. At first sight it would seem that distributions of this type could be disregarded as being physically implausible. However, when they are added to other particular solutions the result can be acceptable. In particular, the true distribution will contain some such invisible distribution whenever its spectrum extends to high frequencies.

In view of the lack of a unique solution to our problem, it may be asked whether any solution can be regarded as the most acceptable representation of the true distribution on the basis of the information available. In the past solutions have been obtained by methods of trial or iteration, using as a criterion of their approximation to the true distribution only that scanning them with the aerial should satisfactorily reproduce the observed curve. If there were only one solution, these methods would lead to it, but we now realize that in radio

astronomy there are many solutions. Unawareness of this fact has led to false conclusions about the distribution of solar and galactic radiation.

If there is no knowledge of $T(\varphi)$ other than that contained in $T_a(\varphi)$, the most that can be done is to restore to their full value those (low frequency)

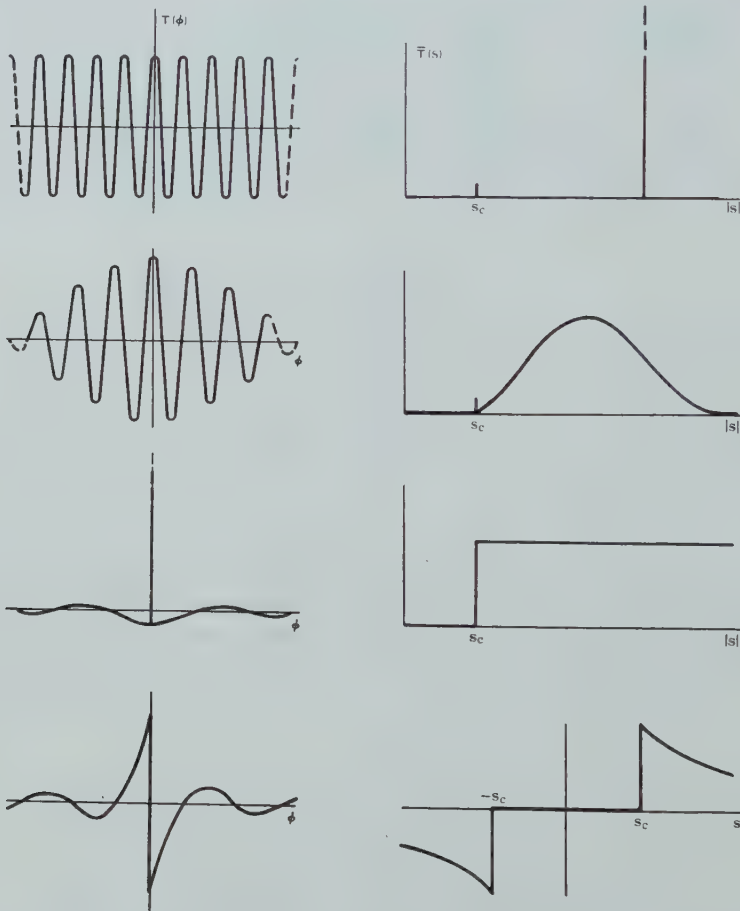


Fig. 7.—Some invisible distributions (left) and their spectra (right).

components of T which, while present in T_a , have been reduced in amplitude. This gives a unique result which we shall call the principal solution, $S(\varphi)$. We may define it as that solution whose spectrum is the same as the spectrum of the true distribution at all values of s for which $\bar{\mathbf{A}}(s) \neq 0$, that is,

$$\left. \begin{aligned} \bar{S}(s) &= \frac{\bar{T}_a(s)}{\bar{\mathbf{A}}(s)}, & (\bar{\mathbf{A}}(s) \neq 0); \\ &= 0, & (\bar{\mathbf{A}}(s) = 0). \end{aligned} \right\} \dots\dots\dots (4)$$

Figure 9 illustrates this for a typical aerial.

When there are no errors, the information contained in \bar{S} is exactly the same as that given by an interferometer consisting of two aerials used at all spacings from zero up to the aerial width. However, the interferometer gives

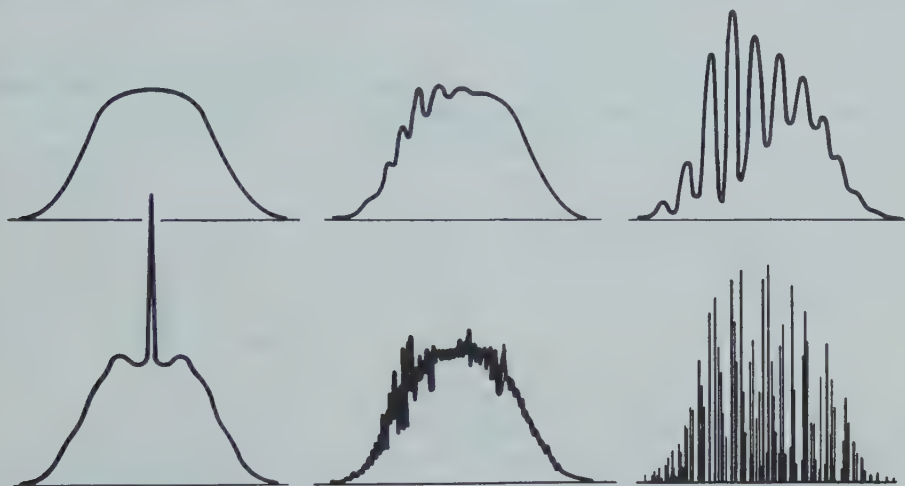


Fig. 8.—A variety of distributions which when scanned with the same aerial all give the same result.

the Fourier components of T directly at their full value, whereas the aerial reduces the components so that they must subsequently be restored. As components

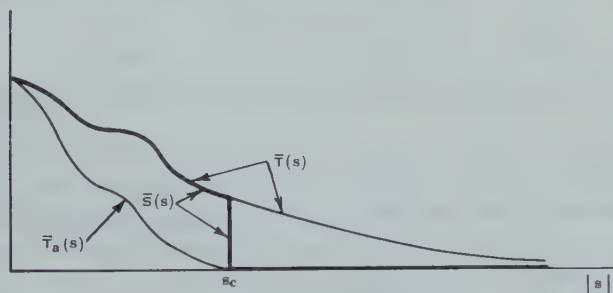


Fig. 9.—The relation between $\bar{S}(s)$ and $\bar{T}(s)$ for an aerial consisting of a uniform aperture.

near the cut-off frequency are very considerably reduced, the corresponding restored values in \bar{S} are subject to very large errors. With this proviso the discussion below of solutions derived from \bar{S} applies equally to the cases of an aerial and of an interferometer with variable spacing.

The degree of approximation of the principal solution to the true distribution depends markedly on the form of T . Thus when T contains no spectral components at those frequencies for which \bar{A} is zero, \bar{S} is identical with T . However, when the spectrum of T is still appreciable at the cut-off frequency s_c , the resulting discontinuity in the spectrum of the principal solution can cause

spurious oscillations in S . The effect of such a discontinuity is illustrated in Figure 10 where the principal solution is shown for two examples of $T(\phi)$, an impulse function and a rectangular distribution. In both cases the principal solution has implausible oscillations, and indeed negative values which, in radio astronomy, are not physically possible.

When the principal solution is not a good approximation to the true distribution, three general methods are available for deriving a physically acceptable distribution which is an approximation to T . In the first approach the discontinuity in \tilde{S} , which is responsible for the spurious oscillations in the

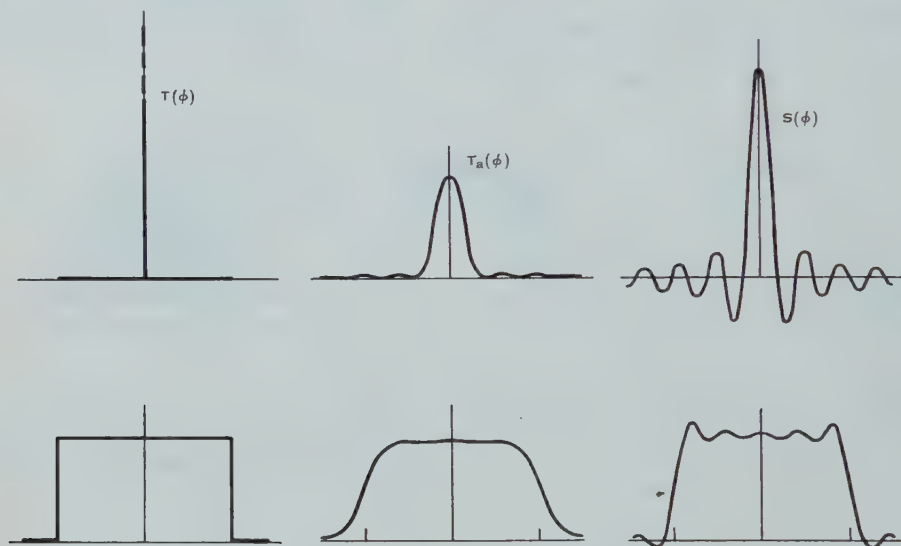


Fig. 10.—Cases where the principal solution $S(\phi)$ is physically impossible.

solution, is changed to a smooth transition by reducing all the components below s_c by some suitable weighting function. This procedure yields a less oscillatory distribution, but at the expense of sharpness. It is commonly used in crystallography (Waser and Schomaker 1953) where the advantage of removing spurious maxima is considered to outweigh the disadvantage of loss of detail. (The information obtained from X-ray or electron diffraction is analogous to that obtained in radio astronomy with a two-aerial interferometer of variable spacing.) In radio astronomy the observed distribution T_a may be regarded as the result of applying such a weighting function, as is illustrated in Figure 9; the method of restoring by successive substitutions discussed in Section VIII is a method of successively reducing the severity of the weighting factor (Fig. 11).

A second approach, also sometimes used in crystallography, is to extrapolate the restored spectrum \tilde{S} to frequencies beyond the cut-off. The method has the advantage that no detail present in \tilde{S} is discarded. However, detail put in by extrapolation must be treated with due caution. In radio astronomy this method has been used to some extent with variable spacing interferometers (O'Brien 1953).

A further method is to construct families of physically plausible distributions and to match their spectra to $\bar{S}(s)$. The form of the trial distributions can be chosen so that the number of parameters to be fitted is as great as the knowledge of $\bar{S}(s)$ allows. This method has the advantage of using not only all the information available in $\bar{S}(s)$, but also any circumstantial knowledge of the true distribution. In practice one may fit either the spectra, or the values obtained at the sampling points when the trial solutions are scanned by the aerial.

The former case arises when variable spacing interferometers are used. An example is furnished by the attempts to find the diameter of the radio stars. Thus Smith (1952) measured $\bar{S}(s)$ for Cassiopeia and Cygnus and expressed his results in terms of the uniform disk whose spectrum agreed best with the observations. This unexceptionable procedure has much to recommend it, especially when, as in this case, a small number of parameters is deemed adequate to describe the measurements. The statement of the result contains a clear indication of the uncertainty.

VIII. RESTORATION BY SUCCESSIVE SUBSTITUTIONS

One of the restoration procedures which has already been used in radio astronomy is equivalent to solving the integral equation of Section III by the method of successive substitutions (see e.g. Lovitt 1950). We study it here for its interest as a current procedure and its importance as a way of finding the principal solution. The process can be understood as follows. Suppose T_{app} is some approximation to T . Then it can be tested by scanning it with the aerial and comparing the result with T_a . The discrepancy $(T_a - \mathbf{A} * T_{app})$ is taken as a first estimate of the difference between T_{app} and T . This leads to a further approximation

$$T_{app} + (T_a - \mathbf{A} * T_{app}). \quad \dots\dots\dots (5)$$

If T_a itself is taken as the initial approximation to T one obtains as the first approximate restoration

$$T_1 = T_a + (T_a - \mathbf{A} * T_a).$$

By applying the same procedure to T_1 we have a second restoration

$$T_2 = T_1 + (T_a - \mathbf{A} * T_1),$$

and the n th restoration is derived from the $(n-1)$ th by the formula

$$T_n = T_{n-1} + (T_a - \mathbf{A} * T_{n-1}).$$

In practice the iteration is halted when smoothing the trial distribution with the aerial gives a result agreeing with T_a within the experimental error.

The approximate solution (5) can be written in two alternative useful forms. In the first form

$$T_n = T_a + \varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_n, \quad \dots\dots\dots (6)$$

where

$$\varepsilon_1 = T_a - \mathbf{A} * T_a = (\delta - \mathbf{A}) * T_a,$$

$$\varepsilon_2 = (\delta - \mathbf{A}) * \varepsilon_1,$$

etc., the successive correction terms being obtained by convolution of the preceding correction with $\delta - \mathbf{A}$. In practice, operation on the correction term obviates the necessity for comparison with T_a as required in the form (5). In the second alternative form the relation is expressed in terms of successive smoothings of the observed distribution :

$$\begin{aligned} T_1 &= 2T_a - \mathbf{A} * T_a, \\ T_2 &= 3T_a - 3\mathbf{A} * T_a + \mathbf{A} * \mathbf{A} * T_a, \\ T_{n-1} &= nT_a - \binom{n}{2} \mathbf{A} * T_a + \binom{n}{3} \mathbf{A} * \mathbf{A} * T_a + \dots + \mathbf{A}^{*(n-1)} * T_a, \quad \dots \quad (7) \end{aligned}$$

where $\mathbf{A}^{*n} = \mathbf{A} * \mathbf{A} * \dots * \mathbf{A}$, the \mathbf{A} appearing n times. van Cittert (1931) obtained the formula in essentially this form.

This elegant procedure, by which one restores smoothed out detail by further smoothing, appears to have been introduced into astronomy by van Cittert (1931) in connexion with the instrumental broadening of spectra. Bolton and Westfold (1950), and (in less explicit form) Hey, Parsons, and Phillips (1948), applied it in radio astronomy to correct their surveys of galactic radiation. It has since been used in other radio-astronomical applications.

Bolton and Westfold thought they had proved that the sequence of successive distributions converged to the true distribution. But, as we have shown, it is in general not possible to recover the true distribution from the observed one. We have therefore studied the process and find (i) that the sequence of distributions does not always converge to a limit and (ii) that when it does the limit is the principal solution defined in Section VII. Whilst developing the proof which follows we had the benefit of a discussion with E. J. Burr.

The Fourier transform of the n th restoration is, by equation (6),

$$\bar{T}_n(s) = \{1 + [1 - \bar{\mathbf{A}}(s)] + [1 - \bar{\mathbf{A}}(s)]^2 + \dots + [1 - \bar{\mathbf{A}}(s)]^n\} \bar{T}_a(s). \quad \dots \quad (8)$$

The series enclosed in braces may be recognized as the first $n+1$ terms of the binomial expansion of $\{1 - [1 - \bar{\mathbf{A}}(s)]\}^{-1}$. Hence, provided that $|1 - \bar{\mathbf{A}}(s)| < 1$,

$$\lim_{n \rightarrow \infty} \bar{T}_n(s) = \frac{\bar{T}_a(s)}{\bar{\mathbf{A}}(s)}.$$

This condition for the convergence of the binomial expansion is not met when $s = s_k$ since $\bar{\mathbf{A}}(s_k) = 0$; but in that case $\bar{T}_n(s)$ is zero for all n since $\bar{T}_a(s)$ is itself zero (equation (8)). Therefore, when $|1 - \bar{\mathbf{A}}(s)| < 1$ for all $s \neq s_k$, $\bar{T}_n(s)$ tends to the spectrum of the principal solution (equation (4)), and T_n tends to the principal solution.

This sufficient condition is met by many aerials. For example, in the case of symmetrical aerials, for which $\bar{\mathbf{A}}(s)$ is real, the condition amounts to requiring $\bar{\mathbf{A}}(s)$ to be non-negative and < 2 . Now $\bar{\mathbf{A}}(s)$ cannot exceed 1 on account of the normalization of $\mathbf{A}(\varphi)$. Hence it is sufficient that a symmetrical aerial have $\bar{\mathbf{A}}(s)$ non-negative in order that restoration by successive substitutions should converge to the principal solution whatever the form of T_a .

It should be noted, however, that $|1 - \bar{\mathbf{A}}(s)| < 1$ for $s \neq s_k$ is a sufficient condition. It is not a necessary condition; consequently convergent results

may be obtained in cases where aerials do not obey it for all values of s , namely, where $\bar{T}_a(s)$ is zero at those values. Should cases of this sort become important, each T_a would have to be examined individually for convergence. The possibility

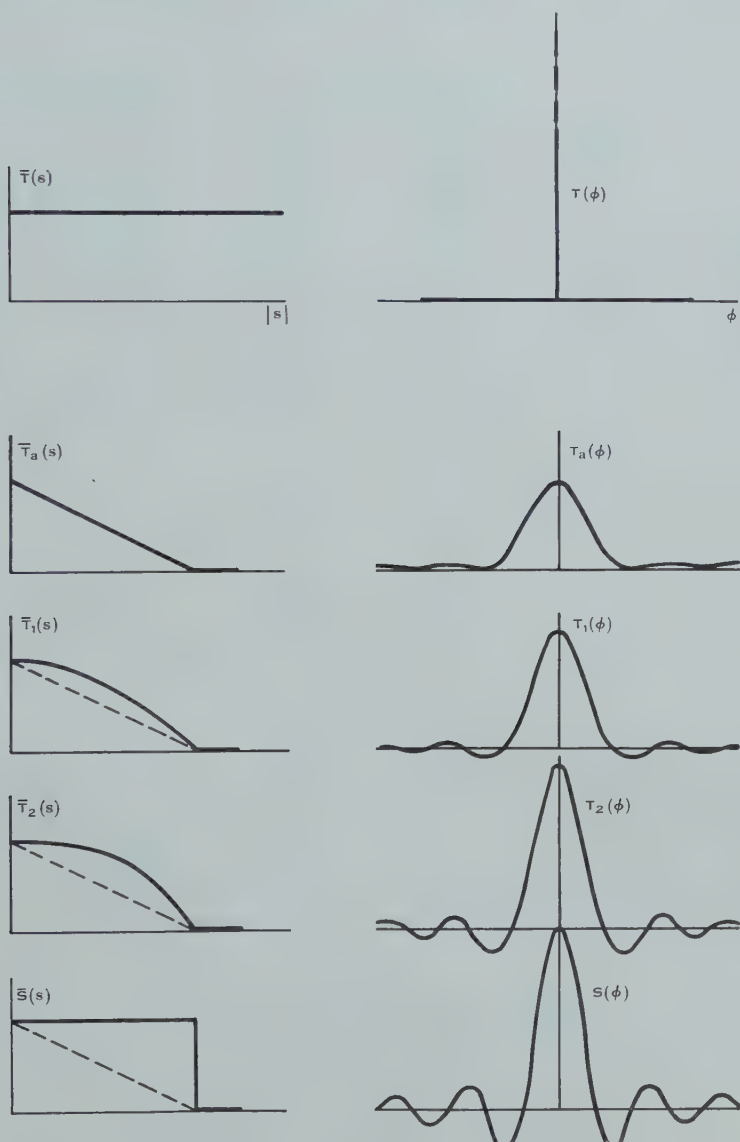


Fig. 11.—Successive restorations (right) of a point source, and their spectra (left).

of T_n approaching S asymptotically, even though ultimately divergent, would also become important.

To summarize, the necessary and sufficient condition for convergence is that $|1 - \bar{\mathbf{A}}(s)| < 1$ for all s such that $\bar{T}_a(s) \neq 0$.

The convergence of $T_n(\varphi)$ to the principal solution is illustrated in Figure 11 for the case of a point source scanned by an aerial consisting of a uniformly illuminated aperture. It is seen that the restorations tend towards a distribution of the form $\varphi^{-1} \sin \varphi$, which was shown above (Fig. 10) to be the principal solution for this case.

For any $T(\varphi)$ the spectra on the left may be interpreted as factors which, when multiplied into $\bar{T}(s)$, give \bar{T} , \bar{T}_a , \bar{T}_1 , \bar{T}_2 , . . . , \bar{S} ; the distributions on the right, on convolution with $T(\varphi)$, give T , T_a , T_1 , T_2 , . . . , S . The distributions on the left may be regarded as weighting functions, in the sense of Section VII; the graphs of T_a , T_1 , and T_2 show that the reduction of the spurious oscillations by drastic weighting is offset by broadening and lowering of the central peak. The distributions on the right may be interpreted as the point source responses of hypothetical aerals, which, if used to survey a temperature distribution, would yield T , T_a , T_1 , etc. as the observed distribution.

Several stages of restoration may be performed in a single smoothing operation. It is evident from the forms (6) and (7) that we may write

$$T_n = R_n * T_a,$$

where the n -fold restoring distribution R_n is given by

$$R_n = \delta + (\delta - \mathbf{A}) + (\delta - \mathbf{A})^{*2} + \dots + (\delta - \mathbf{A})^{*n},$$

or by

$$R_n = (n+1)\delta - \binom{n+1}{2}\mathbf{A} + \binom{n+1}{3}\mathbf{A}^{*2} - \dots + (-1)^n \mathbf{A}^{*n}.$$

For theoretical purposes R_n may be usefully expressed in the closed form

$$R_n = \mathbf{A}_i * [\delta - (\delta - \mathbf{A})^{*(n+1)}],$$

where \mathbf{A}_i , the inverse of \mathbf{A} , is defined by $\mathbf{A}_i * \mathbf{A} = \delta$. Notice that R_n does not in general approach a limit as n tends to infinity. When it is known how many stages of restoration are likely to be justified, it is possible to go direct to the n th restoration by using this formula. However, one should guard against over-restoration which leads simply to an enhancement of errors. In Table 2 we give the restoring distributions R_2 and R_3 for a uniform aperture.

In this case

$$\begin{aligned} A(\Phi) \equiv \mathbf{A}(\Phi) &= \frac{1}{2} \left(\frac{\sin \frac{1}{2} \pi \Phi}{\frac{1}{2} \pi \Phi} \right)^2 \\ &= \begin{cases} 2\pi^{-2} \Phi^{-2}, & \Phi \text{ odd,} \\ \frac{1}{2}, & \Phi \text{ zero,} \\ 0, & \text{other integral } \Phi. \end{cases} \end{aligned}$$

The normalizing factor makes $\Sigma \mathbf{A} = 1$. Table 2 also contains \mathbf{A} , \mathbf{A}^{*2} , and \mathbf{A}^{*3} .

If for any reason T_a contains frequencies beyond the cut-off these will be enhanced $n+1$ times in the n th stage of restoration, as can be shown from the form (7). T_a may contain high frequencies due to errors, or high frequencies may be introduced by initiating the restoration process with a trial solution which contains high frequencies. To avoid the enhancement of such spurious

TABLE 2

THE RESTORING FUNCTIONS R_2 AND R_3 FOR A PLANE APERTURE FED UNIFORMLY AND IN PHASE. THE RESPONSE TO A POINT SOURCE \mathbf{A} AND ITS SUCCESSIVE SELF-CONVOLUTIONS \mathbf{A}^{*2} AND \mathbf{A}^{*3} ARE ALSO TABULATED

$ \Phi $	\mathbf{A}	\mathbf{A}^{*2}	\mathbf{A}^{*3}	R_2	R_3
0	0.5000	0.3333	0.2500	1.8333	2.0833
1	0.2026	0.2026	0.1808	-0.4053	-0.5861
2	0.0000	0.0507	0.0760	0.0507	0.1267
3	0.0225	0.0225	0.0323	-0.0450	-0.0773
4	0.0000	0.0127	0.0190	0.0127	0.0317
5	0.0081	0.0081	0.0120	-0.0162	-0.0282
6	0.0000	0.0056	0.0085	0.0056	0.0141
7	0.0041	0.0041	0.0062	-0.0083	-0.0144
8	0.0000	0.0032	0.0048	0.0032	0.0079
9	0.0025	0.0025	0.0037	-0.0050	-0.0087
10	0.0000	0.0020	0.0030	0.0020	0.0051
11	0.0017	0.0017	0.0025	-0.0033	-0.0058
12	0.0000	0.0014	0.0021	0.0014	0.0035
13	0.0012	0.0012	0.0018	-0.0024	-0.0042
14	0.0000	0.0010	0.0016	0.0010	0.0026
15	0.0009	0.0009	0.0014	-0.0018	-0.0032
16	0.0000	0.0008	0.0012	0.0008	0.0020
17	0.0007	0.0007	0.0011	-0.0014	-0.0025
18	0.0000	0.0006	0.0009	0.0006	0.0016
19	0.0006	0.0006	0.0008	-0.0011	-0.0020
20	0.0000	0.0005	0.0008	0.0005	0.0013
21	0.0005	0.0005	0.0007	-0.0009	-0.0016
22	0.0000	0.0004	0.0006	0.0004	0.0011
23	0.0004	0.0004	0.0006	-0.0008	-0.0014
24	0.0000	0.0004	0.0005	0.0004	0.0009
25	0.0003	0.0003	0.0005	-0.0006	-0.0011
26	0.0000	0.0003	0.0005	0.0003	0.0008
27	0.0003	0.0003	0.0004	-0.0006	-0.0010
28	0.0000	0.0003	0.0004	0.0003	0.0007
29	0.0002	0.0002	0.0004	-0.0005	-0.0008
30	0.0000	0.0002	0.0003	0.0002	0.0006

components, it is therefore necessary to filter the initial data as outlined in Section VI. Then the numerical work should be carried out at the critical interval to prevent the introduction of further high frequencies during the calculations.

IX. APPLICATION TO THE QUIET SUN

(a) Data

In this section the foregoing theory is illustrated for a practical case. The $\mathbf{A}(\varphi)$ is that of the central beam of the high-resolution aerial developed by Christiansen (1953) and is closely of the form $\varphi^{-2} \sin^2 \varphi$. It has a beam width to half power of 3 min of arc and a beam width between zeros of 6.50 min. The peculiar interval $\frac{1}{2}\varphi_c$ is therefore 1.62 min. The angle φ measured in units of this peculiar interval is denoted as before by Φ . Integral values of Φ are

then sufficiently closely spaced to define the band-limited functions. In terms of this normalized angle the aerial response is that given in the second column of Table 2.

For the observed distribution T_a we are indebted to Dr. Christiansen for providing a graph representing the quiet Sun (see the thin curve of Fig. 13). From this graph we have read off T_a at intervals to obtain Table 3. T_a is symmetrical, \bar{T}_a and zero outside the range given. The limits of error quoted were ± 0.020 ($\Phi < 10$) and ± 0.030 ($\Phi > 12$) whilst, on the steep portion $10 < \Phi < 12$, Φ was subject to error up to ± 0.2 .

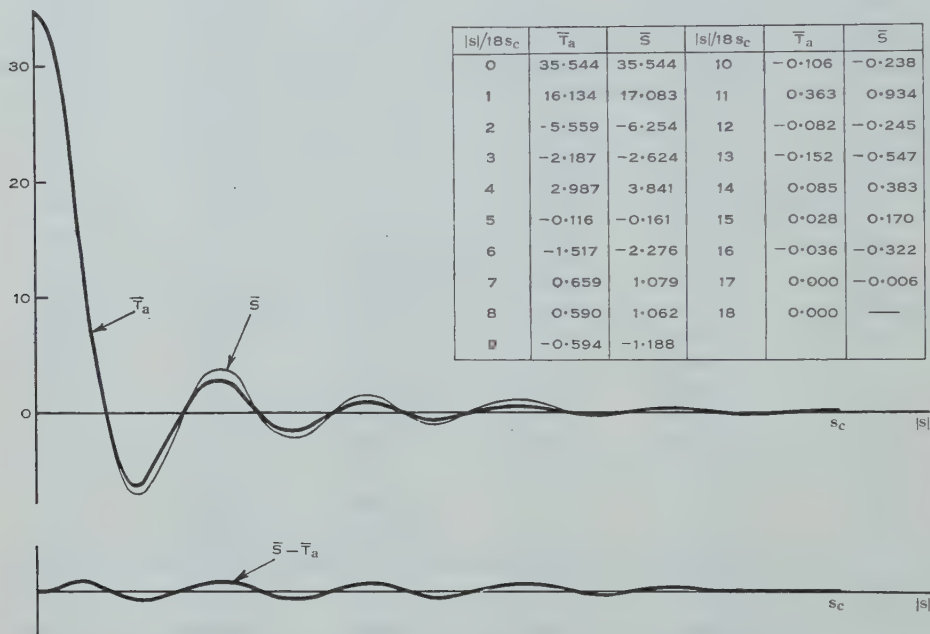


Fig. 12.—Distribution across the quiet Sun (Christiansen and Warburton 1953). The spectra of the observed distribution \bar{T}_a and the principal solution \bar{S} (above) and the difference between them (below).

The data T_a should be free from Fourier components of period shorter than φ_c , but the reduction processes intervening between the actual records and the adopted T_a , especially the subtraction of sunspot effects, might well introduce high frequencies. In the present case it was found that no correction for high frequencies needed to be made. In the following work we therefore adopt for our observed distribution the band-limited function defined by $T_a(\Phi)$, with Φ integral.

(b) Restoration and the Spectrum

We now attempt restoration by using the Fourier transforms of the two functions \mathbf{A} and T_a . The Fourier transform of $\mathbf{A}(\varphi)$ is the triangular function

$$\bar{\mathbf{A}}(s) = 1 - |s/s_c|, \quad |s| < s_c, \\ = 0, \quad |s| > s_c.$$

The transform of T_a must be computed numerically. By our previous results, we may compute instead the transform of $\text{III}(\Phi)T_a$, and the Fourier integral then reduces *without approximation* to a summation of products. For each value of s there are 18 products as only 18 of the sequence of numbers defining T_a are non-zero. The transform of T_a , shown by the heavy curve in Figure 12, is oscillatory and decays to zero amplitude at the cut-off frequency s_c . (For reference the calculated values have been tabulated on the figure.)

TABLE 3

THE BRIGHTNESS DISTRIBUTION ACROSS THE QUIET SUN AT 21 CM (CHRISTIANSEN AND WARBURTON 1953). THE OBSERVED DISTRIBUTION T_a AND THE PRINCIPAL SOLUTION S

Φ	T_a	S	Φ	T_a	S	Φ	T_a	S
0	1.000	1.018		1.005		13	0.115	0.084
	1.000		7	1.005	1.024		0.087	
1	1.001	1.019		1.005		14	0.064	0.027
	1.001		8	1.004	1.036		0.047	
2	1.001	1.020		1.004		15	0.034	0.012
	1.002		9	1.002	1.089		0.024	
3	1.002	1.022		0.966		16	0.015	-0.007
	1.003		10	0.854	0.995		0.008	
4	1.003	1.024		0.680		17	0.004	-0.013
	1.003		11	0.490	0.440		0.000	
5	1.004	1.029		0.324		18	0.000	-0.014
	1.004		12	0.212	0.120		0.000	
6	1.004	1.028		0.153		19	0.000	-0.011

From \bar{T}_a and $\bar{\mathbf{A}}$, \bar{S} may be calculated by division. It is shown as the light curve in Figure 12. The difference curve $\bar{S} - \bar{T}_a$ shows the spectral distribution of the correction applied. The amplitudes of the Fourier components near the cut-off frequency are small, even after restoration.

To obtain the principal solution S it is now necessary to transform \bar{S} . Unfortunately the distribution S is not of finite extent, so that values of \bar{S} at discrete intervals do not fully define \bar{S} . It is therefore necessary to know \bar{S} at sufficiently small intervals of s to permit evaluation of the transform of \bar{S} by numerical integration. On the other hand the Fourier integral need be evaluated at discrete intervals only. However, the labour involved in establishing \bar{S} and transforming it is formidable. For interest we show by the heavy line in Figure 13 the result of transforming $\text{III}(18s/s_c)\bar{S}(s)$, that is, substituting for the Fourier integral of \bar{S} a summation based on the 18 spot values of \bar{S} given on Figure 12. The oscillations A, B, C prove to be spurious and arise from the use of too coarse an interval of s .

(c) Successive Substitutions

Restoration of T_a by the method of successive substitutions is illustrated in Figure 14. The figure shows T_a , with the limits of uncertainty indicated by

shading; $A*T_a$; the successive corrections $\varepsilon_1, \varepsilon_2, \varepsilon_3$; and the corresponding restorations T_1, T_2, T_3 . All the smoothing processes were computed using discrete values of the variables. From the three successive stages of restoration which are illustrated it may be seen that the corrections rapidly drop below the significant level. If T_3 is tested by forming $A*T_3$, it is found to differ from T_a by ε_4 which is so small that $A*T_3$, if plotted to the scale of Figure 14, would barely be distinguishable from T_a .

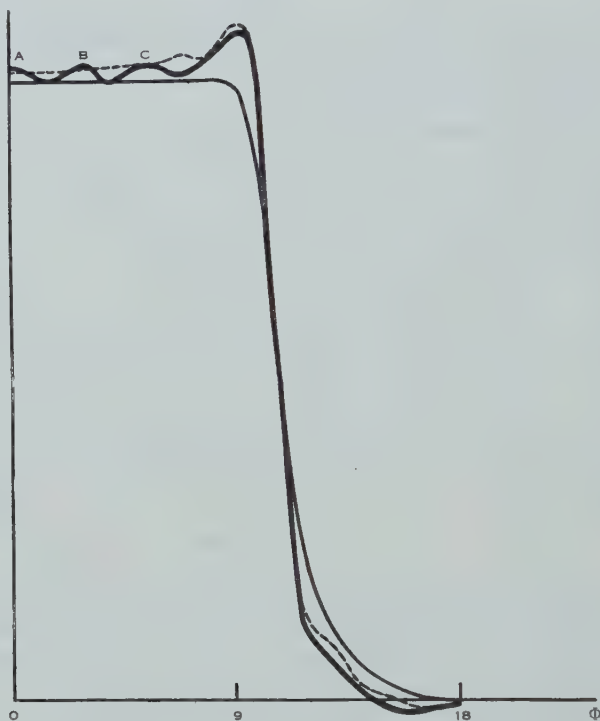


Fig. 13.—The observation distribution T_a (thin curve), the principal solution S (dotted curve), and an approximation to the principal solution obtained by restoring the transform of T_a (heavy curve).

It is not necessary in practice to obtain all the curves plotted in Figure 14, which was constructed simply to illustrate the process. Using the method outlined in Section VIII, T_3 may be obtained directly from T_a in one step.

In order to ascertain the principal solution with some accuracy, for the purpose of checking different proposals for approximating to it, we have carried the process to the 11th stage. The distribution so obtained is labelled S in Table 3 and is plotted as a dotted curve in Figure 13. One may verify that it is in fact the principal solution by scanning it with the aerial pattern. The result agrees with T_a to within one unit in the last decimal place at all integral values of Φ . The values of S between the tabulated values are to be understood

as those given by the method of interpolation described in Section VI; consequently S is free from frequencies beyond s_c . The order of accuracy of this determination of S is of course much better than the 20 or 30 units in the last decimal which the accuracy of the observations would justify.

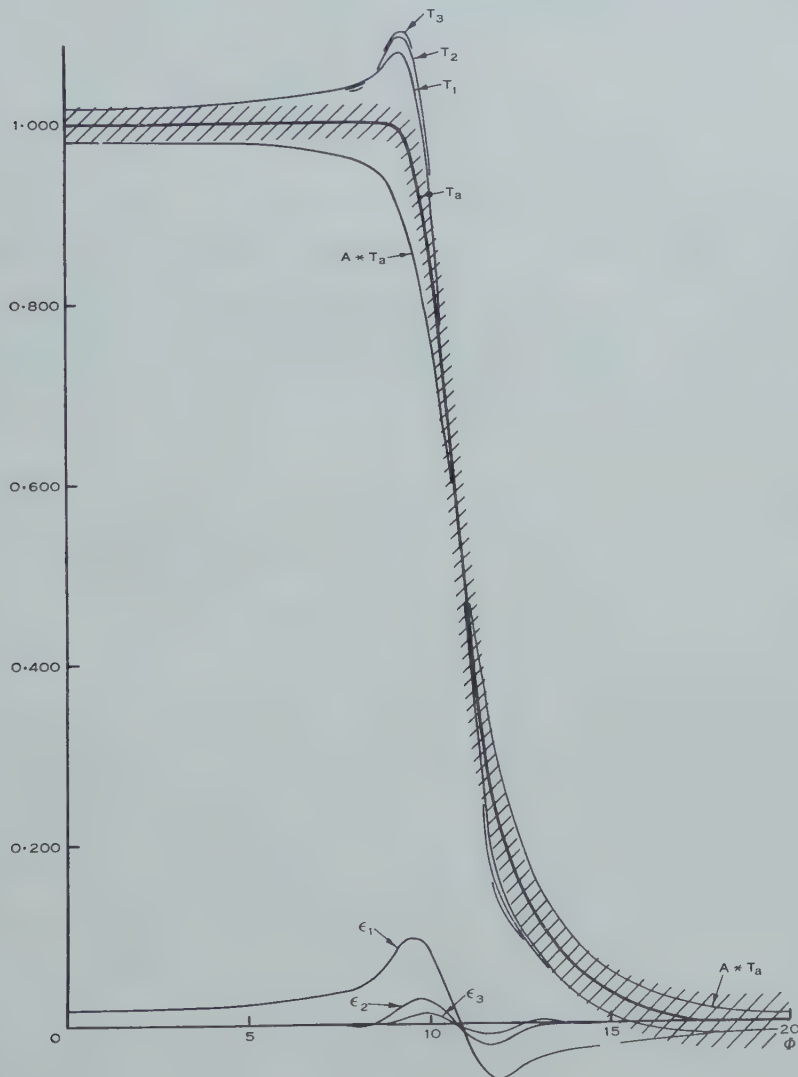


Fig. 14.—The observed distribution T_a , the successive corrections ϵ_1 , ϵ_2 , ϵ_3 , and the successive restorations T_1 , T_2 , T_3 .

X. REFERENCES

- BOLTON, J. G., and WESTFOLD, K. C. (1950).—Galactic radiation at radio frequencies. I. 100 Mc/s survey. *Aust. J. Sci. Res. A* **3**: 19.
- BOOKER, H. G., and CLEMMOW, P. C. (1950).—The concept of an angular spectrum of plane waves, and its relation to that of polar diagram and aperture distribution. *Proc. Instn. Elect. Engrs. III* **97**: 11.

- BOOKER, H. G., RATCLIFFE, J. A., and SHINN, D. H. (1950).—Diffraction from an irregular screen with applications to ionospheric problems. *Phil. Trans.* **242**: 579.
- BROWN, R. H., and HAZARD, C. (1953).—A model of the radio-frequency radiation from the Galaxy. *Phil. Mag.* **44**: 939.
- BURGER, H. C., and VAN CITTERT, P. H. (1932).—Wahre und scheinbare Intensitätsverteilung in Spektrallinien. *Z. Phys.* **79**: 722.
- BURR, E. J. (1955).—Sharpening of observational data in two dimensions. *Aust. J. Phys.* **8** (in press).
- CHRISTIANSEN, W. N. (1953).—A high-resolution aerial for radio astronomy. *Nature* **171**: 831.
- CHRISTIANSEN, W. N., and WARBURTON, J. A. (1953).—The distribution of radio brightness over the solar disk at a wavelength of 21 centimetres. II. The quiet Sun—one-dimensional observations. *Aust. J. Phys.* **6**: 262.
- VAN CITTERT, P. H. (1931).—Zum Einfluss der Spaltbreite auf die Intensitätsverteilung in Spektrallinien. *Z. Phys.* **69**: 298.
- DOETSCH, G. (1937).—"Theorie und Anwendung der Laplacetransformation." (Springer: Berlin.)
- FELLGETT, P. B., and SCHMEIDLER, F. B. (1952).—On the sharpening of observational data with special application to the darkening of the solar limb. *Mon. Not. R. Astr. Soc.* **112**: 445.
- HEAVISIDE, O. (1922).—"Electromagnetic Theory." Vol. 2, p. 116. (Benn: London.)
- HEY, J. S., PARSONS, S. J., and PHILLIPS, J. W. (1948).—An investigation of galactic radiation in the radio spectrum. *Proc. Roy. Soc. A* **192**: 425.
- VAN DE HULST, H. C. (1941).—The determination of the true profile of a spectral line. *B.A.N.* **9**: 225.
- VAN DE HULST, H. C. (1946).—Generalization of some methods for solving an integral equation of the first kind. *B.A.N.* **10**: 75.
- LOVITT, W. V. (1950).—"Linear Integral Equations." (Dover: New York.)
- O'BRIEN, P. A. (1953).—The distribution of radiation across the solar disk at metre wavelengths. *Mon. Not. R. Astr. Soc.* **113**: 597.
- PATERSON, M. S. (1950).—Calculation of the correction for instrumental broadening in X-ray diffraction lines. *Proc. Phys. Soc. A* **63**: 477.
- PIDDINGTON, J. H. (1951).—The origin of galactic radio-frequency radiation. *Mon. Not. R. Astr. Soc.* **111**: 45.
- SHANNON, C. E. (1949).—Communication in the presence of noise. *Proc. Inst. Radio Engrs.* **37**: 10.
- SMERD, S. F. (1950).—Radio-frequency radiation from the quiet Sun. *Aust. J. Sci. Res. A* **3**: 34.
- SMITH, F. G. (1952).—The measurement of the angular diameter of radio stars. *Proc. Phys. Soc. Lond. B* **65**: 971.
- SNEDDON, I. N. (1951).—"Fourier Transforms." (McGraw-Hill: New York.)
- STOKES, A. R. (1948).—A numerical Fourier-analysis method for the correction of widths and shapes of lines on X-ray powder photographs. *Proc. Phys. Soc.* **61**: 382.
- TRUMPLER, R. J., and WEAVER, H. F. (1953).—"Statistical Astronomy." (Univ. California Press.)
- WASER, J., and SCHOMAKER, V. (1953).—Fourier inversion of diffraction data. *Rev. Mod. Phys.* **25**: 671.
- WOODWARD, P. M., and LAWSON, J. D. (1948).—The theoretical precision with which an arbitrary radiation-pattern may be obtained from a source of finite size. *J. Instr. Elect. Engrs.* **III 95**: 363.

THE ATTENUATION OF LIGHT BY METEORIC DUST IN THE UPPER ATMOSPHERE

By R. G. GIOVANELLI*

[*Manuscript received March 2, 1954*]

Summary

An examination of Zacharov's results on the wavelength variation of attenuation of solar radiation following the Perseid meteor shower shows that the effects could be produced either by absorbing particles of diameter about 10^{-5} cm or less, or by transparent particles of diameter in the range 5×10^{-5} to 10^{-4} cm. The rapid disappearance of attenuation, however, can be explained only if the particles evaporate on falling and it is concluded that they are ice crystals formed on nuclei of meteoric origin at a height of about 80 km where there is a temperature minimum. From estimates of the light scattered by these ice crystals it is deduced that they would be visible as noctilucent clouds.

I. INTRODUCTION

The recent claim by Bowen (1953) of a correlation between days of heavy rainfall and meteor showers stimulates interest in the nature of meteoric dust in the high atmosphere.

Zacharov (1952) has recently analysed measurements of atmospheric transparency made over the period 1908–1920 by Abbot, Fowle, and Aldrich (1913, 1922) using the so-called “long” method. He found a reduction of transparency whose maximum value, of the order of 2 per cent., occurs some 3 days after the maximum of the Perseid meteor shower, and he considered that the recovery was not complete for some 24 days. At maximum, the continuous decrease in attenuation with increasing wavelength was taken to indicate particles of diameter not exceeding 10^{-5} cm. The time of recovery in attenuation was taken to indicate a duration of atmospheric pollution lasting no more than about 24 days; a time of fall of this order requires particles of some 10^{-3} cm diameter, from which it was deduced that agglomeration must occur, probably at a level of about 80 km.

We show here, on the assumption of spherical particles, that the attenuation of sunlight associated with the Perseid shower is due to ice crystals of average diameter about 7×10^{-5} cm, at an elevation of about 80 km. Disappearance of attenuation is due to evaporation of the particles after they fall through the region of temperature minimum. The particles should scatter sufficient light to be detectable in directions near the Sun, provided the atmosphere be sufficiently pure, and they should be visible as a noctilucent cloud around dusk; Bowen (1953) has in fact found that noctilucent clouds are always associated with meteor showers.

* Division of Physics, C.S.I.R.O., University Grounds, Sydney.

II. ATTENUATION OF LIGHT BY SMALL SPHERICAL PARTICLES

(a) Conductors

The scattering and attenuation of a light beam by small spherical particles was first discussed by Mie (1908). For very small absorbing particles, $2a \ll \lambda$, the scattering cross section Q_s is given by

$$Q_s = \frac{2^7 \pi^5 a^6}{3 \lambda^4} \left| \frac{N^2 - 1}{N^2 + 2} \right|^2 = \pi a^2 \left[\left(\frac{2a}{\lambda} \right)^4 \cdot \frac{2^3 \pi^4}{3} \left| \frac{N^2 - 1}{N^2 + 2} \right|^2 \right], \quad \dots (2.1)$$

where $N = k_1/k_2$ is the ratio of the propagation constants in the particle and in the surrounding medium, a is the particle radius, and λ the wavelength in air.

The cross section for attenuation, Q_t , is given by

$$Q_t = \frac{2^3 \pi^2 a^3}{\lambda} \cdot \operatorname{Re} \left(-i \frac{N^2 - 1}{N^2 + 2} \right) = \pi a^2 \left[\left(\frac{2a}{\lambda} \right)^2 2\pi \operatorname{Re} \left(-i \frac{N^2 - 1}{N^2 + 2} \right) \right]. \quad \dots (2.2)$$

Since $N^2 = (2\pi\mu/\lambda)^2(1 + i\kappa)^2$ where μ is the refractive index and κ the absorption index, the cross sections may be computed from the optical constants of the particle. Typical values of K_s and K_t , where $Q = K(\pi a^2)$, are given in Table 1 for several metals at 0.59μ and for carbon.

TABLE 1
TYPICAL VALUES OF K_s AND K_t FOR VERY SMALL ABSORBING SPHERES
($2a \ll \lambda$)

Material	K_s	K_t
Silver	$420 \times (2a/\lambda)^4$	$0.39 \times (2a/\lambda)$
Iron	340	4.7
Nickel	390	4.8
Carbon	92	2.6

An upper limit beyond the range of validity of the above expressions is set by the conditions $Q_s \leq Q_t$. For the three metals considered in Table 1 and for carbon, expressions (2.1) and (2.2) are certainly no longer valid when the diameter is greater than about $\lambda/4$.

Ruedy (1941, 1942) has calculated the attenuation and scattering by carbon spheres over a larger range of diameters, $2a \lesssim \lambda$. Values of K_t derived from his graphs are given in Table 2. For large values of diameter, K_t decreases to a value of 2. Ruedy's results showed K_s to be very small compared with K_t until $2a/\lambda \approx 0.125$, then to rise rapidly to about $0.3K_t$ for $2a/\lambda \approx 0.25$ and then more slowly to about $0.5K_t$ when $2a/\lambda \approx 0.75$ to 1.25 . For large diameters, $K_s = 0.5K_t$, as readily follows, e.g. from van de Hulst's (1946) discussion.

van de Hulst has also given results for the attenuation and scattering of light by various absorbing media. For iron at 0.42μ , K_t rises almost linearly to about 2.8 when $2a/\lambda = 0.25$, then slowly to a maximum of 3.0 when $2a/\lambda \approx 0.5$, from which it slowly drops to 2 for large diameters. The behaviour is rather like that of carbon. For metals the scattered light is intermediate between that for transparent media and carbon.

TABLE 2
 K_t FOR CARBON AND WATER SPHERES

$\frac{2a}{\lambda}$	K_t	
	Carbon	Water
0.25	1.9	<0.15
0.5	2.6	0.33
0.75	3.0	1.0
1.0	2.5	1.8
1.5		3.4
2.0		3.9
3.0		2.2

(b) *Insulators*

For insulators, the attenuation cross section is equal to that for scattering, and for very small spheres is proportional to $(2a/\lambda)^4$. Data have been calculated by Stratton and Houghton (1931), Ruedy (1943a), and la Mer (1943), and are summarized for water in Table 2. On the whole, the results for different refractive indices are rather similar, the cross sections being compressed or expanded along the $2a/\lambda$ axis. The values of $2a/\lambda$ for maximum values of K_t are given in Table 3.

TABLE 3
 VARIATION WITH REFRACTIVE INDEX OF THE SPHERE
 DIAMETER FOR MAXIMUM K_t

Refractive Index	$2a/\lambda$ for K_t (max.)
2.25	2.5
1.33	1.9
1.44	1.5
1.50	1.3
1.55	1.2

III. THE DIAMETER OF PARTICLES IN THE UPPER ATMOSPHERE AS DERIVED FROM SELECTIVE ATTENUATION

Zacharov was able to obtain the variation of atmospheric transparency t with wavelength at the time of maximum absorption, and data from his graphs are given in Table 4.

Although the changes Δt in transparency are small, the uniform variation with wavelength suggests that the effect is real. Zacharov inferred from the selectivity that the particle diameters did not exceed 10^{-5} cm. It is possible, however, to interpret these results otherwise, as we shall show.

If the particles be conducting, then for diameters less than about $\lambda/4$ the attenuation may be derived from (2.2). It is usually agreed that this results in the attenuation varying approximately as λ^{-1} , although the term involving

the propagation constant does depend on wavelength. For the longer wavelengths, the values of $\Delta t/t$ in Table 4 vary approximately in this manner. The trend of the curve suggests a maximum at a wavelength somewhat shorter than 0.35μ . By comparison with the value of $2a/\lambda$ for maximum K_i with carbon (see Table 2), we infer that for $\lambda = 0.35 \mu$, $2a/\lambda \lesssim 0.5$, i.e. the diameters of the particles, if conducting, are of the order of 1.7×10^{-5} cm or less.

TABLE 4
OBSERVED ATTENUATION AS A FUNCTION OF WAVELENGTH (ZACHAROV 1952)

Wavelength (μ)	Transmittance t	Δt	$\frac{\Delta t}{t}$
0.35	0.60	0.0245	0.041
0.4	0.72	0.025	0.035
0.45	0.80	0.031	0.039
0.5	0.85	0.0225	0.026
0.6	(0.89)*	0.020	(0.022)
0.7	0.93	0.014	0.015
0.8	(0.95)	0.012	(0.012)
1.0	(0.96)	0.0095	(0.0095)
1.2	(0.97)	0.0075	(0.0075)
1.6	(0.98)	0.0025	(0.0025)

* Values in parentheses are estimates.

On the other hand, suppose the particles are transparent; then a fair fit of the values of K_i to $\Delta t/t$ can be achieved by taking the maximum of the K_i curve to be at about 0.35μ , as can be seen by comparing the cross sections for water, Table 2, with $\Delta t/t$, Table 4. If the refractive index is 1.33 (water droplets) the particle diameter is about 7×10^{-5} . van de Hulst (1949) has given reasons to suppose that the refractive index of particles in interstellar space may be about 1.25, while the index of most minerals is appreciably higher. In view of the uncertainties, including those of the experimental data, it would seem that diameters in the range 5×10^{-5} to 10^{-4} cm are likely.

It is clear that selective attenuation can be explained by either absorbing or transparent particles, but the diameters required differ in order of magnitude.

IV. THE TIME VARIATION OF ATTENUATION

In order to explain the disappearance of attenuation, Zacharov suggested that the particles agglomerate and fall to the ground over a period of 24 days. This time, however, is somewhat uncertain. Zacharov plotted against time the 3-day averages of the transmission, the central 3-day group in each year being centred on the maximum of the Perseid shower. For each wavelength, the mean curve, derived from 10 years' observations, shows fluctuations about the mean; and, from inspection, these do not seem asymmetrical more than 24 days after the time of maximum attenuation. But the attenuation 3 days after the maximum of the shower shows the only increase which is statistically significant,

being of the order of three times the standard deviation, and the time for recovery is thus uncertain.

The attenuation coefficient of n particles of diameter $2a$ in a column of unit cross section is

$$k = K_t \pi n a^2. \quad \dots\dots\dots (4.1)$$

If dust particles grow in size without coalescing, then the attenuation varies as $K_t a^2$, and increases continuously for all types of material. This clearly cannot be the process of disappearance of attenuation.

Since the total volume V of all particles is $4\pi n a^3/3$, then

$$\frac{k}{V} = \frac{3K_t}{4a}. \quad \dots\dots\dots (4.2)$$

If particles coalesce, so that a increases while V remains constant, their attenuation varies as K_t/a . For both absorbing and transparent particles of initial diameters as found in the preceding section, K_t/a decreases with a , and the attenuation varies roughly as a^{-1} , so that in each case coalescence would reduce the attenuation.

Zacharov has plotted times for particles of the density of water to fall through various heights in the atmosphere. To fall to the ground in 24 days the particles would require diameters of the order of 10^{-3} cm.; for shorter times of fall the diameters would need to be even greater. However, the attenuation produced by transparent particles after coalescence to a diameter of 10^{-3} cm would be reduced by a factor of about 0.05–0.1; for absorbing particles the factor would be even smaller, about 0.02. This argument would suggest that the disappearance of attenuation is due to coalescence rather than to the fall of the particles. We shall now consider whether or not coalescence is physically possible.

V. PROCESSES AFFECTING PARTICLE DIAMETER

For particles not affected by condensation or evaporation there is only one conceivable mechanism of particle growth, coalescence by collision. Kunkel (1948) has estimated the collision rate between a particle of radius a_1 , charge q_1 , and particles of radius a_2 , charge q_2 , and concentration N falling freely in air of viscosity η . If

$$|C_1| > |V_0| (a_1 + a_2)^2, \quad \dots\dots\dots (5.1)$$

where

$$C_1 = \frac{q_1 q_2 (a_1 + a_2)}{6\pi\eta a_1 a_2},$$

$$V_0 = \frac{2g\rho(a_1^2 - a_2^2)}{9\eta},$$

g being the acceleration due to gravity and ρ the density of a particle, then Kunkel's discussion shows that the mean time τ between collisions is

$$\tau = -\frac{1}{4\pi N C_1}. \quad \dots\dots\dots (5.2)$$

During day-time, particles in the upper atmosphere will be positively charged due to photoelectric ionization, and collisions will be improbable. At night-time, recombination will result in particles whose average charge is zero, although the average absolute magnitude of the charge per particle will differ from zero, being such that

$$\frac{1}{2} \frac{q^2}{a} \approx kT. \quad \dots\dots\dots (5.3)$$

If $T=250^\circ\text{K.}$, $a=3.5 \times 10^{-5} \text{ cm}$, then $q \approx 1.6 \times 10^{-9} \text{ e.s.u.}$

The value of N is obtained by noting that at the time of maximum attenuation $a \approx 3.5 \times 10^{-5} \text{ cm}$, $K_i \approx 2$ for transparent particles, so from (4.1),

$$n \approx 3 \times 10^6 \text{ cm}^{-2}.$$

If these are assumed to be in a layer 10 km thick, then $N \approx 3 \text{ cm}^{-3}$, which is much greater than Bowen's estimate of 10^{-6} cm^{-3} for the concentration of meteoric dust particles in the upper atmosphere. Then with $\eta \approx 1.7 \times 10^{-4} \text{ g cm sec}^2$, $a_1 \approx a_2 \approx 3.5 \times 10^{-5} \text{ cm}$, $a_1^2 - a_2^2 \leq 10^{-9} \text{ cm}^2$, $q_1 \approx -q_2 \approx 2 \times 10^{-9} \text{ e.s.u.}$, and $\rho = 1 \text{ g cm}^{-3}$, it follows that (5.1) applies, and $\tau \approx 10^9 \text{ sec}$, which is so long that such collisions may be completely neglected. Agglomeration cannot occur, and the attenuation cannot have been due to meteoric dust alone.

It seems very likely, however, that the particles consist of nuclei of meteoric origin, on which water vapour has condensed to form water droplets or ice crystals. Since this requires relative humidities of the order of 100 per cent., the regions in which condensation can occur are limited. Although we have scanty knowledge of the variation of water vapour content of the atmosphere with altitude, the approximate temperature and pressure distributions are known (Havens, Koll, and la Gow 1952). Above 40 km, the only region where the total pressure exceeds the saturated water vapour pressure is in the neighbourhood of 80 km, where the temperature has a minimum value of about 190°K. and this is the only region where condensation can occur. Furthermore, at such low temperatures any condensation must be as ice crystals. Zacharov has estimated the velocity of fall of spherical particles of unit density and diameter $7 \times 10^{-5} \text{ cm}$ to be of the order of 10 cm sec^{-1} at the 80 km level; the lifetime of the ice crystal, which is the time to fall through the low temperature region, about 10 km thick, is thus of the order of a day.

VI. BRIGHTNESS OF PARTICLES

(a) These particles must cause an increase in the brightness of the sky near the Sun. The distribution of light scattered in the forward direction may be obtained approximately on the assumption that it results from diffraction by opaque disks, in which case it readily follows that the change b in sky luminance at an angle θ to the Sun, produced by a layer having n particles in a column of unit section along the line of sight, is

$$b = \frac{nB\Omega a^2}{\theta^2} J_1 \left(\frac{2\pi a\theta}{\lambda} \right), \quad \dots\dots\dots (6.1)$$

B and Ω being the mean luminance of and the solid angle subtended by the Sun. But

$$J_1(x) = \frac{x}{2} - \frac{x^3}{16} + \frac{x^5}{384} - \dots,$$

so

$$\frac{1}{\theta} J_1\left(\frac{2\pi a\theta}{\lambda}\right) = \frac{\pi a}{\lambda} - \frac{1}{2}\left(\frac{\pi a}{\lambda}\right)^3 \theta^2 + \dots$$

and varies but slowly with θ while θ is small, falling to $1/\sqrt{2}$ of its maximum value when $\theta \simeq 0.5\lambda/2a$, i.e. 0.25 if $\lambda/a = 1$. Provided θ is small in this sense,

$$\frac{b}{B} = \frac{\pi^2 n a^4 \Omega}{\lambda^2} = \frac{\pi k a^2 \Omega}{K_t \lambda^2} \dots \dots \dots (6.2)$$

With $k = 0.025$, $K_t \approx 2$ and $2a/\lambda = 1.25$, then

$$\frac{b}{B} \approx 10^{-6}.$$

This is measurable, and of the order of magnitude of the luminance of a very pure sky near the Sun, but, in view of the variable nature of the light scattered in the low atmosphere, considerable care would be needed to detect variations of this amount. There is, in fact, only very limited published data on the brightness of the sky near the Sun, the most extensive tables available being of the Fraunhofer Institut (1951-53) which give brightness of the sky a few minutes from the Sun's limb at wavelength 0.5303μ . At such small angles of scattering the brightness and its fluctuations are so high as to hide increases of the above order of magnitude without many more observations, and these should be made preferably not too close to the Sun's limb.

(b) The amount of light scattered at large angles from an ice crystal depends on the shape and size of the particle, and on the angle of scattering. A short extrapolation from a graph given by Ruedy (1943*b*) indicates that at right angles to an incident beam of illumination e the intensity of a spherical particle, refractive index 1.33 and $2a/\lambda = 1.25$, is about $8 \times 10^{-2} e a^2$. A column of such particles, n per unit section along the line of sight, will have a luminance of $8 \times 10^{-2} n e a^2$; if $n = 3 \times 10^6 \text{ cm}^{-2}$, $a = 3.5 \times 10^{-5} \text{ cm}$ and $e \approx 10^4 \text{ lumen ft}^{-2}$ (as in sunlight), then the luminance is about 3 cd ft^{-2} or 9 ft lamberts . It is unlikely that such a change in sky brightness could be measured during daylight, or be seen visually even if it possessed a structure. At dusk, however, when the background sky luminance becomes very small the cloud of ice crystals should be readily seen. Bowen's (1953) findings that noctilucent clouds have been reported only at times of meteor showers thus lead to the conclusion that Zacharov's absorption is probably due to noctilucent clouds. There would seem to be no measurements of the luminance of such clouds, but it must be of the above order of magnitude.

VII. CONCLUSIONS AND DISCUSSIONS

The above results may be summarized briefly as follows. From the variation in absorption with wavelength, it is deduced that attenuation of sunlight within

a few days of the Perseid meteor shower is due either to absorbing particles of diameter of 10^{-5} cm or less, or to transparent particles of diameter 5×10^{-5} to 10^{-4} cm. The rapid disappearance of absorption can be due only to decrease in size of the particles and it is concluded that they are ice crystals or water droplets condensed on nuclei of meteoric origin. The only region where the relative humidity can be high enough for this is around 80 km, where because of the low temperature (190°K) the particles must be ice. It is then deduced that the mean diameter is 7×10^{-5} cm, with about 3×10^6 particles/cm² along the line of sight, and of the order of 3 particles/cm³.

Particles of the above size and concentration would be visible as a noctilucent cloud around dusk, though not bright enough to be seen in daytime. In the neighbourhood of the Sun, however, there should be an increase in sky brightness of the order of 10^{-6} of that of the Sun, which should be detectable provided suitable precautions are taken against scattering by particles in the low atmosphere.

VIII. ACKNOWLEDGMENT

The author is grateful to Dr. D. F. Martyn for valuable suggestions relating to this work.

IX. REFERENCES

- ABBOT, C. G., FOWLE, F. E., and ALDRICH, L. B. (1913).—*Ann. Astrophys. Obs. Smithson. Instn.* Vol. 3.
 ABBOT, C. G., FOWLE, F. E., and ALDRICH, L. B. (1922).—*Ann. Astrophys. Obs. Smithson. Instn.* Vol. 4.
 BOWEN, E. G. (1953).—*Aust. J. Phys.* **6**: 490.
 FRAUNHOFER INSTITUT (1951–53).—*Sonnen-Zirkulare*.
 HAVENS, R. J., KOLL, R. T., and LA GOW, H. E. (1952).—*J. Geophys. Res.* **57**: 59.
 VAN DE HULST (1946).—*Rech. Astr. Obs. Utrecht* Vol. 11, Part 1.
 VAN DE HULST (1949).—*Rech. Astr. Obs. Utrecht* Vol. 11, Part 2.
 KUNKEL, W. (1948).—*J. Appl. Phys.* **19**: 1053.
 LA MER (1943).—O.S.R.D. Rep. No. 1857.
 MIE, G. (1908).—*Ann. Phys., Lpz.* **25**: 377.
 RUEDY, R. (1941).—*Canad. J. Res. A* **19**: 117.
 RUEDY, R. (1942).—*Canad. J. Res. A* **20**: 25.
 RUEDY, R. (1943a).—*Canad. J. Res. A* **21**: 79.
 RUEDY, R. (1943b).—*Canad. J. Res. A* **21**: 99.
 STRATTON and HOUGHTON (1931).—*Phys. Rev.* **38**: 159.
 ZACHAROV, I. (1952).—*Bull. Cent. Astr. Inst. Csl.* **3**: 82.

SHORT COMMUNICATIONS

SEA SURFACE TEMPERATURES*

By F. K. BALL†

Introduction

The surface of the sea is losing heat by evaporation and by long-wave radiation exchange with the sky, both of these rates being of the order of 10^{-2} W cm $^{-2}$ in clear weather. Heat lost in this way must be provided by conduction upward from the water beneath and downward from the air above. Short-wave radiation need not be considered since it is not absorbed at the surface. It seems possible therefore that on days of light wind the "skin" temperature of the sea might be appreciably less than the temperature of the layers beneath. Now sea surface temperatures are usually measured by means of a dip bucket and it is clear that water entering the bucket is derived in varying amounts from various depths below the surface, so that the temperature of the mixture will not generally be equal to the skin temperature of the sea. In view of these considerations an experiment was carried out with the object of determining the skin temperature by measuring the long-wave radiation emitted from the sea and comparing it with the dip bucket temperature.

Method

The observations were taken from the end of Mordialloc pier during the late summer and autumn of 1953. The pier is approximately 300 yd long and the depth of the water at the end is about 16 ft. All measurements were made in clear weather with light winds, partly to obtain steady conditions and partly to reduce the number of variables in the experiment. Furthermore if a cool skin does exist it will reach its maximum development in these circumstances. A Linke Feussner actinometer placed about 6 ft above the water was used to measure the radiation. The instrument was inclined at about 10° to the downward vertical so that the supporting timbers of the pier were not in its field of view. The galvanometer used in conjunction with the actinometer was housed in a large box with a slot at one end for viewing the scale. Despite all precautions taken to reduce the effect of vibration caused by sea and wind it was at times not possible to take reliable measurements. Considerable fluctuation in reading could also be caused by specular reflection of sunlight from the oscillating sea surface.

In steady conditions the galvanometer can be read to the nearest 0.5 mm, which corresponds to a temperature difference between instrument and sea of

* Manuscript received June 15, 1954.

† Section of Meteorological Physics, C.S.I.R.O., Melbourne.

about 0.15°C . This small deflexion is produced by a temperature difference between the thermojunctions of the actinometer of the order of 0.01°C . In spite of its thick copper casing there are almost always temperature gradients within the instrument which can produce much bigger temperature differences and correspondingly larger spurious deflexions. If the spurious deflexion produced in this way were to remain constant during the course of an observation then it would be of no importance. However, the temperature field within the instrument is unfortunately never quite steady, especially when the instrument is in use in the open air, and the spurious or "zero" deflexion may change by several millimetres in the 30 sec necessary to take an observation. This means that it is necessary to take an observation several times until one is fortunate enough to obtain one during which the zero deflexion changes only by 1 mm or less. Deflexions can usually only be obtained to the nearest millimetre because of this, and consequently temperature differences can only be measured to about 0.3°C . This inaccuracy is inherent in the instrument itself and cannot be reduced by increasing the sensitivity of the galvanometer.

The Linke Feussner measurements of long-wave radiation from the sea surface do not give its temperature directly, firstly because the sea is not a perfect black body and secondly because the air between instrument and sea absorbs and emits long-wave radiation. It is possible to allow for these effects in the following manner. Suppose the reflection coefficient of the sea is α and the transmissivity of the air between instrument and sea is $1-\beta$. If the temperature of the sea surface is T_1 , the incoming radiation from the sky is R , and a representative temperature of the air is T_2 then the apparent temperature of the surface as measured with the Linke Feussner actinometer, T say, is given by

$$\sigma T^4 = [(1-\alpha)\sigma T_1^4 + \alpha R][1-\beta] + \beta\sigma T_2^4, \dots\dots\dots (1)$$

where σ is the Stefan-Boltzmann constant. Now R is approximately equal to $4\sigma T_1^4/5$, so

$$T^4 \approx T_1^4 \left(1 - \frac{\alpha}{5}\right) - \beta(T_1^4 - T_2^4),$$

therefore

$$T \approx T_1 \left(1 - \frac{\alpha}{20}\right) - \beta(T_1 - T_2). \dots\dots\dots (2)$$

The coefficient β depends on the vapour pressure of the air and can be estimated from an Assman psychrometer reading taken at the level of the instrument. T_2 was put equal to the dry-bulb temperature at instrument height, the temperature gradient being accounted for by taking only 170 cm air path instead of 200 cm. The reflection coefficient of water at normal incidence is given by Brunt (1939, p. 121, Fig. 35) for wavelengths up to 18μ . The average value weighted according to Planck's law for a temperature of 300°K , is about 0.03. This does not contradict the value of 0.06 usually quoted in the literature since this was calculated from Fresnel's formulae (Ångström 1915) on the assumption that the refractive index is $4/3$. It is a weighted mean over all angles of incidence taking account of variation of both reflection coefficient

and quantity of incident radiation. At angles of incidence from 0° to 30° Fresnel's formulae give 0.02 for the reflection coefficient. It will be assumed here that the reflection coefficient at an angle of incidence of 10° is 0.03.

Results

The data obtained are shown in Table 1. The differences between the actual surface temperature and the dip bucket temperature have a mean value of -0.25°C . The standard deviation is 0.32°C , which is consistent with the accuracy of the observations. Since there are 12 observations the difference is significant at the 5 per cent. level. On only one occasion (February 19) was there an appreciable warm skin and that was also the only occasion when the air was appreciably warmer than the water. These results are in substantial agreement with those of Bruch (1940). The experiment was carried out in conditions favourable for the development of a cool skin, i.e. conditions of clear sky and light wind. In general therefore the temperature difference will be less than 0.25°C but in very favourable circumstances may be higher. It is unlikely that any serious errors will be made if surface temperatures are quoted to the nearest 1°F when measured with dip buckets suitably designed to minimize other sources of error.

TABLE 1
RESULTS OF SEA SURFACE TEMPERATURE MEASUREMENTS

Date	Time	$T_2 - T_1$ ($^\circ\text{C}$)	$\beta \times 10^2$	δ^* (mb)	ΔT^\dagger ($^\circ\text{C}$)
1953					
Feb. 19	1530	+4.15	13.5	10.0	+0.35
" 26	1412	-2.0	15	9.7	-0.2
" "	1427	-1.65	15	9.5	-0.4
" "	1536	-2.25	15	9.3	+0.15
Mar. 4	2140	-4.1	13	9.3	0.0
" "	2146	-4.2	13	9.5	-0.2
Apr. 7	1305	-2.1	10.5	9.7	-0.45
" "	1407	-2.0	11	9.6	-0.15
" 10	1508	-1.3	14	5.1	-0.75
" 15	1457	+0.5	16.5	3.3	-0.45
" "	1507	+0.5	16.5	3.3	-0.55
May 29	1520	-3.1	11.5	3.5	-0.5

* δ is the difference between the saturation vapour pressure at the sea surface and the vapour pressure at instrument height.

† ΔT is the difference between surface temperature and dip bucket temperature.

Stability of the Cool Skin

If the rate of loss of heat from the surface is $10^{-2} \text{ W cm}^{-2}$ then the temperature gradient in the subsurface laminar layer is $1.7^\circ\text{C cm}^{-1}$. This will produce a temperature difference of 0.25°C in 0.15 cm which is therefore the approximate

thickness of the laminar layer. Rayleigh (1916) has shown that it is possible for a fluid to remain "unstably" stratified provided the following inequality holds

$$\frac{\rho_1 - \rho_0}{\rho_0} < \frac{27\pi^4 k \nu}{4g h^3},$$

where k is the diffusibility for heat and ν is the kinematic viscosity, ρ_1 the density at the top and ρ_0 the density at the bottom of the layer whose thickness is h . In the present case the condition becomes

$$\Delta T < 13^\circ \text{C}.$$

Thus on this basis at any rate there is no reason to doubt the existence of a laminar layer.

References

- ÅNGSTRÖM, A. (1915).—Smithson. Misc. Coll. Vol. 65, No. 3.
 BRUCH, H. (1940).—Veröff. Inst. für Meeresk. Univ. Berl., A, Heft 38.
 BRUNT, SIR DAVID (1939).—"Physical and Dynamical Meteorology." 2nd Ed. (Cambridge Univ. Press.)
 RAYLEIGH, LORD (1916).—*Phil. Mag.* **32**: 529-46.

MICROINTERFEROMETRIC EXAMINATION OF NUCLEAR EMULSION PLATES*

By B. S. THORNTON†

The photographic emulsion is one of several means of detecting charged particles emitted in nuclear reactions or from radioactive materials, and emulsion technique has developed enormously over the past few years; it is now an invaluable laboratory tool. Whilst the principles of nuclear emulsion microscopy are well established, the application of interference microscopy to the study of nuclear emulsions is new. Although it seems that the interferometric technique would have no real advantage at the present time over the conventional microscopic methods of track examination, it is felt desirable to report some observations on the examination of tracks by interferometry and to suggest some lines of possible investigation.

Nuclear tracks are delineated by grains of developed silver oriented in the dry gelatin layer after photographic processing, and the resulting heights and depths in the emulsion are clearly revealed by the interference microscope. One disadvantage of the method when used in reflection is that it is only applicable to tracks very near the surface of the emulsion where the developed silver grains are slightly exposed. This limits the method to those experiments where tracks enter the emulsion at near grazing incidence. However, if a

* Manuscript received June 28, 1954.

† Division of Metrology, C.S.I.R.O., University Grounds, Sydney.

MICROINTERFEROMETRIC EXAMINATION OF NUCLEAR EMULSION PLATES

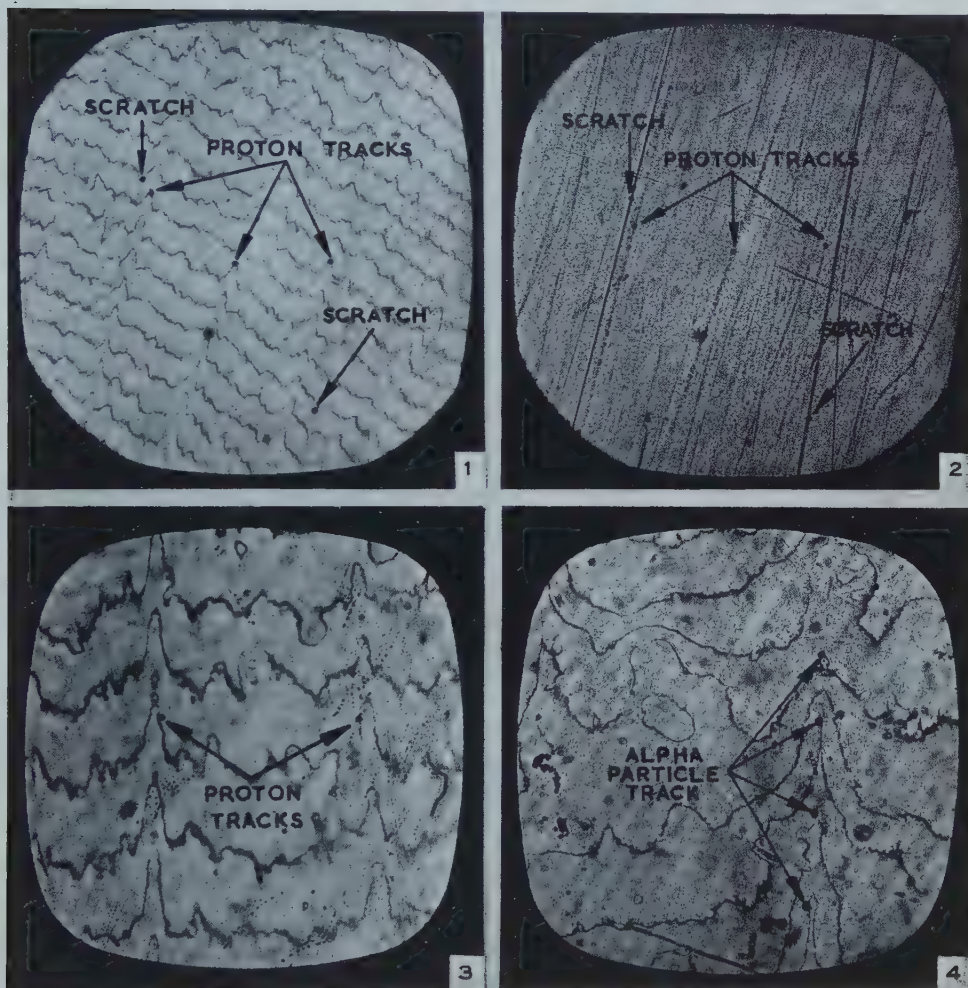


Fig. 1.—Interferogram of proton tracks. Fringes dip opposite ways for tracks and scratches.

Fig. 2.—Proton tracks in emulsion corresponding to Figure 1. (The dark lines are scratches.)

Fig. 3.—Proton tracks in C2 emulsion.

Fig. 4.— α -Particle track and local distortion in D1 emulsion.

MICROINTERFEROMETRIC EXAMINATION OF NUCLEAR EMULSION PLATES

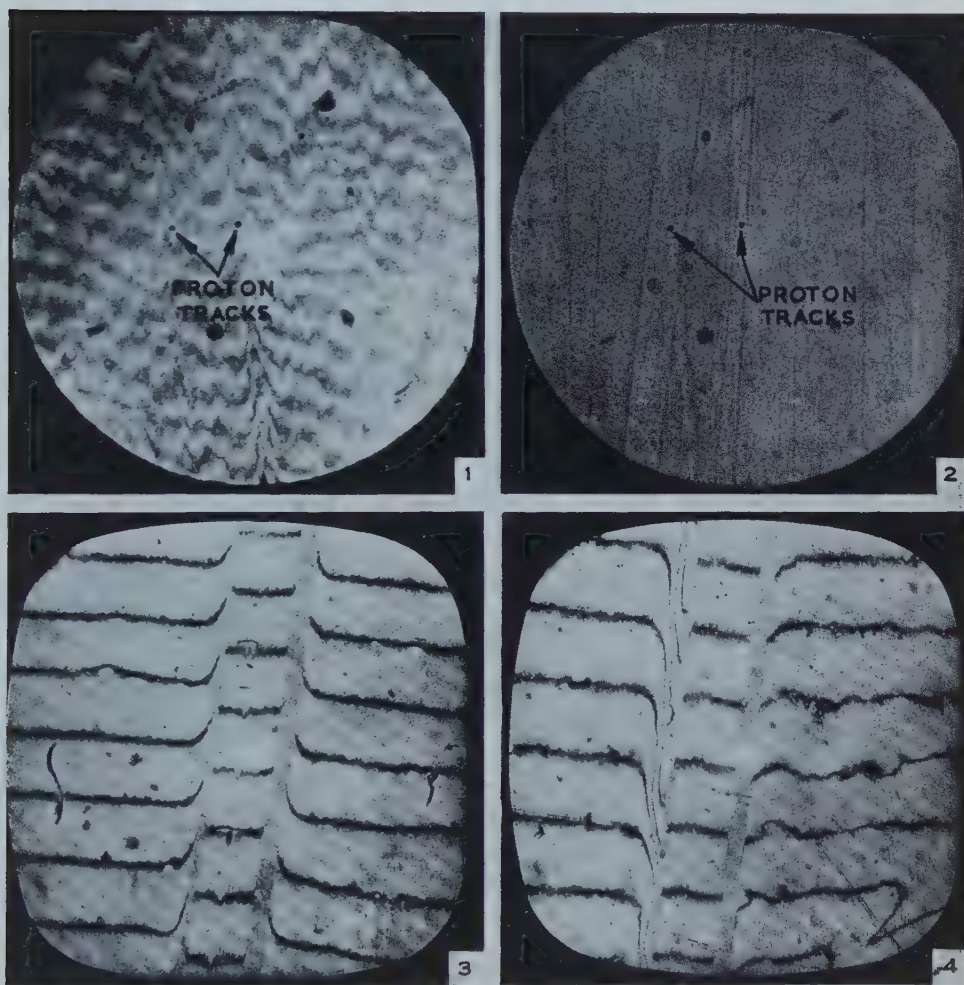


Fig. 1.—Two-beam interferograms of proton tracks.

Fig. 2.—Proton tracks corresponding to Figure 1.

Fig. 3.—Interferogram of groove ruled in unprocessed emulsion. Increasing fringe order from bottom to top.

Fig. 4.—Interferogram of groove ruled in plate after processing. Increasing fringe order from top to bottom.

section of emulsion stripped from the glass is examined in a transmission interference system using equal chromatic order fringes, the paths of tracks well below the surface of the emulsion could be examined. Preliminary experiments have indicated that care must be taken in mounting the stripped emulsion so that it is quite flat and free from warping over the region to be examined.

Typical reflection microinterferograms ($\times 51$ in each case) of proton tracks in a lithium loaded Ilford C2 emulsion, and α -particle tracks in a boron loaded Ilford D1 emulsion, are shown in Plate 1, Figures 1-4, and Plate 2, Figures 1-2, together with some corresponding microphotographs also taken in reflection. The fact that the proton tracks in the microphotographs of Plate 1, Figure 2, and Plate 2, Figure 2, appear black and white respectively is due only to a slight out of focus effect. Using a Linnik interference microscope much higher surface magnifications can be obtained if required. Apart from distortion, there seems to be a small difference between the types of interferograms from α -tracks and from proton tracks. The α -tracks examined seem to give much sharper and longer peaks in the fringe pattern. The plates were vacuum coated with silver to 80 per cent. reflectivity. Actually silvering is not strictly necessary, since two-beam fringes can be obtained from the unsilvered plate as shown in Figure 1 of Plate 2.

The tracks usually have certain dip angles which can be measured in the conventional manner of change of focus and, if the angle of dip is known, the grain size along the track can be found from the microinterferogram. This could possibly serve to help distinguish between different types of particles, especially those of high ionizing power where "clogging" is observed and the tracks form a continuous line, making it impossible to measure density of ionization by grain counting. If the plates receive grain-gradation development (Van der Grinten 1948) in order to increase the differentiating power, then the interference technique combined with this method should be capable of easily differentiating between particles and simultaneously providing the range-energy curve, since the size of the grain agglomerations is given by the fringes and the mean size of a single developed grain is approximately 0.3μ . The method of grain gradation does not seem to be widely accepted but has on several occasions given very useful information (Yagoda 1949).

A possible improvement in the accuracy of measurement of the angle of dip of nuclear tracks would be introduced by adapting the principle of the optical coincidence gauge (Gardner and Case 1931) to the measurements of change of depth. The most obvious instrument for making these measurements is the ordinary microscope with a fine focusing micrometer attachment as is used in conventional nuclear research microscopes. However, there is a lack of precision arising from depth of focus unless an objective of large numerical aperture is used, in which case the working distance becomes very small. The optical coincidence microscope described by Gardner and Case is similar in principle to a range-finder and the precision is intermediate between that of a screw micrometer and an interferometer.

Multiple beam interferometry or even two-beam interferometry can be used to obtain an accurate measure of the shrinkage factor of an emulsion and its variation with humidity and other factors, if part of the plate is left unprocessed and a groove ruled across the plate through the treated and untreated emulsion. A typical multiple beam microinterferogram of the groove in the unprocessed emulsion is shown in Plate 2, Figure 3, and that for the processed emulsion in Plate 2, Figure 4. The groove was ruled with a ruling tool 0.008 in. wide in this case. The shrinkage factor can be calculated by measuring the groove depths from observations of the fringe fractions in three different wavelengths. Small changes in the shrinkage factor can be determined quite readily in this way. Although the high accuracy thus obtained is not warranted in conventional practice, it may become of importance when more precise measurements are required.

Thanks are due to Professor E. W. Titterton and Mr. T. A. Brinkley of the Research School of Physical Sciences, Australian National University, for supplying the plates and for their helpful discussions and criticism. The author would also like to thank Mr. N. A. Esserman, Chief of the Division of Metrology, and Mr. C. F. Bruce, of the Division, for their suggestions, interest, and encouragement.

References

- GARDNER, I. C., and CASE, T. A. (1931).—*Bur. Stand. J. Res.* **6**: 229.
VAN DER GRINTEN, W. (1948).—*Bull. Amer. Phys. Soc.* **23** (3): 51.
YAGODA, H. (1949).—"Radioactive Measurements with Nuclear Emulsions." pp. 285 and 295. (John Wiley & Sons: New York.)

THE NOISE GENERATED IN A COIL WITH A FERROMAGNETIC CORE*

By G. BUILDER† and D. HANEMAN‡

It is well known that Barkhausen noise will be induced in a coil which has a ferromagnetic core, if the core is subjected to any considerable degree of varying or alternating magnetization due to current flowing in the coil or due to a magnetizing field from some external source. This effect is ascribed to irreversible magnetization processes, such as irreversible domain wall movements, and has been investigated in considerable detail. It has recently been utilized to estimate the amount of the contribution of such irreversible processes to the total magnetization (Tebble, Skidmore, and Corner 1950).

* Manuscript received September 1, 1954.

† School of Physics, University of Sydney.

‡ School of Physics, University of Sydney; present address, Division of Radiophysics, C.S.I.R.O., University Grounds, Sydney.

It is also well known that the insertion of a non-magnetic conducting core into a coil will increase the thermal noise generated in the coil, and that the total noise depends on the effective coil resistance, with the core present, in accordance with Nyquist's formula. This increase in noise can be interpreted on the one hand as due to thermal fluctuations in the core and, on the other hand, as corresponding to the losses caused by eddy currents in the core when an alternating test voltage is applied to the coil terminals to measure its resistance.

Furthermore, if such a conducting core were also endowed with ideal magnetic susceptibility (i.e. if it were a homogeneous isotropic material displaying a perfectly linear relation between magnetizing field and magnetization) the consequent increase in magnetic flux and eddy currents associated with a current through the coil, or associated with thermal fluctuations, would result in an increase in coil resistance and a corresponding increase in the thermal noise generated in the coil. It is to be emphasized that this effect is due solely to the increase in flux caused by the ideal susceptibility of the core.

Any real ferromagnetic core must also cause such an increase in thermal noise on account of its conductivity and its finite magnetic susceptibility. This must be true even if the core is not subject to any varying or alternating magnetization which would give rise to Barkhausen noise effects. It is not, however, immediately obvious

- (a) whether the real ferromagnetic core will contribute still further to the noise because of the departure of its actual properties from the ideal, i.e. because of the non-linearity of its characteristics, or because of its domain structure;
- (b) whether it is permissible, in spite of the actual ferromagnetic properties of the core, to use Nyquist's formula to calculate the total noise from the measured value of the coil resistance.

It is therefore necessary to consider the conditions under which Nyquist's formula is applicable, and the relevant properties of ferromagnetic materials.

Nyquist's formula.—The mean square value $E_f^2 df$ of the noise e.m.f. generated in the frequency range df , in an electrical circuit of which the impedance has a resistive component R_f at frequency f , is given approximately by

$$E_f^2 df = 4kTR_f df,$$

where k is Boltzmann's constant and T the absolute temperature of the circuit. Nyquist derived this formula for linear electrical circuits by simple thermodynamic considerations.

To review the conditions under which this formula is applicable, reference may well be made to the definitions used by Callen and Welton (1951) in their derivation of a generalization of it. The formula can be regarded as a special case of a general relation between the fluctuations of a "generalized force" (i.e. voltage), in a linear system in equilibrium, and a parameter which characterizes a dissipative process (i.e. electrical resistance) in the system. A system is said to be dissipative if it is capable of absorbing energy when subjected

to a time-periodic perturbation (e.g. alternating test voltage); it is said to be linear if the power dissipation is quadratic in the magnitude of the perturbation.

These definitions emphasize that Nyquist's formula can be used whatever the physical nature of the dissipative processes that determine the value of the electrical resistance; it is only necessary that the system be linear in the sense defined, for the range of conditions of interest, and this corresponds to the common use of the term in reference to electrical circuits.

The fact that ferromagnetic materials exhibit marked non-linearity in many of their practical uses is not necessarily relevant; the linearity of the system under consideration must be investigated specifically, in accordance with the above concepts, under the particular conditions of interest. It is to be noticed that even electrical circuits constructed of non-magnetic materials are linear only under restricted conditions (e.g. metallic conductors at constant temperature) and, on the other hand, that the inclusion in an electrical system of an essentially non-linear device such as a thermionic tube does not necessarily cause the system to exhibit non-linearity if the conditions considered are sufficiently restricted.

The relevant characteristics of ferromagnetic materials.—The first significant investigation of the behaviour of ferromagnetic materials for very small changes in magnetization was that made by Lord Rayleigh (1887). He found that the susceptibility was constant for various samples of iron and steel for magnetizing-field changes up to about 2×10^{-2} oersteds and down to the smallest changes, about 2×10^{-5} oersteds, for which measurements could be made. Many subsequent investigations have confirmed this linearity of the relation between magnetizing field and magnetization for small variations of the magnetizing field, whether there is a steady polarizing field present or not, and irrespective of the magnetic history of the specimen. For our present purposes, a sufficient summary of this work, and of the present views on the magnetization processes involved, is given in a recent paper by Tebble and Corner (1950).

The greater part of the magnetization process is ascribed to irreversible changes commonly associated with the Barkhausen effect; but a substantial part of the process is due to reversible changes such as continuous movements of domain boundaries through successive equilibrium conditions. These changes are not reversible in the thermodynamic sense only because the conductivity of actual core materials permits more or less localized eddy currents to be set up by the changes, with consequent dissipation.

The reversibility referred to does not in itself imply linearity in each of the individual local changes which contribute to the magnetization; but for a core material consisting of a complex polycrystalline aggregate it is to be expected, and it is found, that the total of the individual reversible contributions is linearly related to the magnetizing field.

The term reversible susceptibility is used for the susceptibility of the material to changes in the applied magnetizing field which are so small that Barkhausen effects are not observed. Figure 1, taken from the paper by Tebble and Corner, shows the relation of the measured values of reversible susceptibility χ_r to the

magnitude ΔH of the applied alternating magnetizing field, for a specimen of hard-drawn iron. This curve is typical of conditions at the foot of the initial magnetization curve or at values of the steady magnetizing field of the order of the coercive force, where the reversible susceptibility is greatest and the extent of the reversible regions is least. It is seen that, in this case, the measured values of the reversibility are almost constant for values of ΔH up to 10^{-2} oersteds. It is clear from this and similar evidence that there is a linear relation between magnetizing field and magnetization for small variations or alternations in the field.

There does not appear to be any direct experimental evidence relating the eddy current losses associated with this reversible magnetization to the amplitude

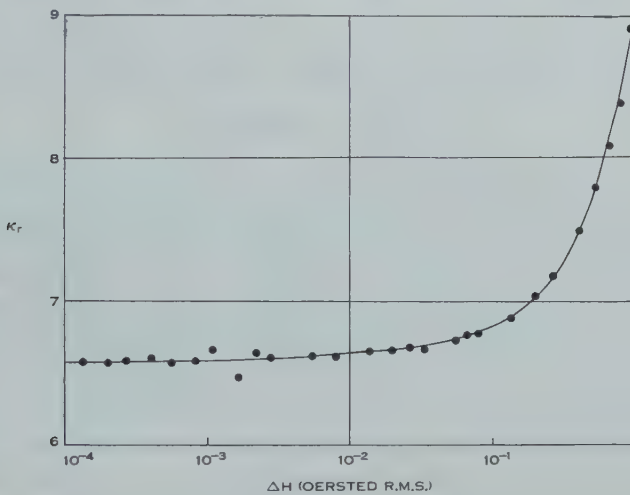


Fig. 1.—Effect of magnitude ΔH of the applied alternating field on the measured reversible susceptibility of hard-drawn iron.

of the applied alternating field ; but it is a reasonable inference from present knowledge of magnetization processes that, in polycrystalline core materials, these losses will increase quadratically with the field amplitudes within the regions in which the susceptibility is constant.

The noise generated in a coil with a ferromagnetic core.—When there is no varying or alternating magnetization of the core, the noise can therefore be calculated using the Nyquist formula, provided that any alternating test voltage applied to the coil to measure its resistance R_f is chosen so that any consequent alternating magnetizing field applied to the core is so small that the magnetization process is linear.

The measured value of the coil resistance R_f will include the effects of the resistivity of the coil windings and of the electrical resistivity of the core, as well as the effect of the constant reversible susceptibility of the core. The increase in noise corresponding to the susceptibility of the core may be interpreted as the result of spontaneous fluctuations in the equilibrium magnetization of the core,

or it may be interpreted as the result of increased magnetic flux associated with thermal fluctuation currents.

This investigation was carried out in the School of Physics in the course of work supported by the Research Grants Committee of the University of Sydney.

References

- CALLEN, H. B., and WELTON, T. A. (1951).—*Phys. Rev.* **83**: 34.
 RAYLEIGH, LORD (1887).—*Phil. Mag.* (5) **23**: 224.
 TEBBLE, R. S., and CORNER, W. D. (1950).—*Proc. Phys. Soc. Lond. B* **63**: 1005.
 TEBBLE, R. S., SKIDMORE, I. C., and CORNER, W. D. (1950).—*Proc. Phys. Soc. Lond. A* **63**: 739.

SCATTERING OF 4.9 MEV PROTONS BY ^{27}Al †

By K. B. MATHER‡

Since 1934 there have been 24 determinations of the level scheme of ^{27}Al . However, even the more recent findings are far from harmonious and the question naturally arises as to how much of the apparent discord is due to target contamination, inadequate resolution, etc., and how much follows from the use of different nuclear reactions where, presumably, selection rules may operate to control transition probabilities. For instance, four of the more recent reports claimed a doublet at ~ 1 MeV while the remainder found only a single level.

The present work contains a new determination of the ^{27}Al levels below 4 MeV, obtained by scattering 4.90 MeV protons from a thin aluminium target. Scattered particles were recorded at the 45° position on Ilford C2 plates inclined at 3° to the direction of scatter. The scattering chamber has been described elsewhere (Mather 1951). Track ranges were measured and are shown in Figure 1. Eighty-nine per cent. of the tracks are associated with the elastic scattering peak (1) centred at $160\ \mu$, while 11 per cent. had ranges less than $130\ \mu$ and were interpreted as due to inelastic scattering. Their positions correspond to the Q values listed in Table 1.

The width due to straggling of 4–5 MeV protons at half maximum is ~ 4 per cent. (e.g. Dyer 1952), i.e. ~ 160 keV, which is therefore approximately the limit of resolution achievable with a photographic emulsion at this energy. The observed half width of peak (1) is 175 keV, the difference being due to incident proton beam width and target thickness. The evidence for two groups at $\sim 110\ \mu$ comes from the observed width (~ 300 keV) and the asymmetry of the group. The envelope can be fitted by assuming two levels at 1.01 and ~ 0.83 MeV respectively, the latter having, perhaps, 40 per cent. of the intensity of the former. The position of the lesser group is uncertain to the extent of ± 0.1 MeV. The closeness of these levels will prevent them from being cleanly

† Manuscript received September 3, 1954.

‡ Australian Atomic Energy Commission, University of Melbourne.

resolved by range measurements in emulsions, which probably accounts for the failure of many observers to detect two levels in this region.

Table 1 also lists intensity estimates for each group based on the relative number of tracks (and calling the total number of charged particle products 100 per cent.). The final column gives absolute values of differential cross

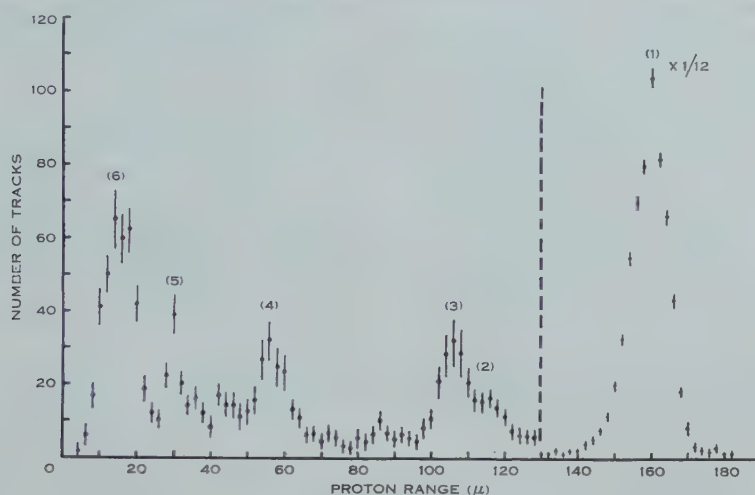


Fig. 1.—Range spectrum of 4.9 MeV protons scattered from a thin aluminium target. The elastic scattering peak (1) has been scaled down 12 times for convenient plotting, change of scale occurring at 130 μ .

sections in the laboratory system at 45° based on the measured track density per unit area of plate and the various constants of irradiation.

TABLE 1

Q VALUES FOR $^{27}\text{Al}(p,p')^{27}\text{Al}$ AND INTENSITIES OF ELASTIC AND INELASTIC GROUPS AT SCATTERING ANGLE 45°
Excited states = — Q values

Group	Q (MeV)	Intensity Estimate (%)	$d\sigma/d\omega$ (45°) $\times 10^{26}$ (cm ² sterad ⁻¹)
(1) Elastic ..	0 (ground)	89.1	6.2
(2) Inelastic ..	—0.83	0.8 (?)	0.056 (?)
(3) „ ..	—1.01	2.0	0.14
(4) „ ..	—2.22	1.7	0.12
(5) „ ..	—3.01	1.2	0.084
(6) Unresolved protons and α -particles			

The background between (4) and (5) is severe and may be due to an unresolved group corresponding to an excited state between 2.2 and 3 MeV. Reilley *et al.* (1952) reported a level at 2.78 MeV.

Group (6) is certainly not single and the impression of a peak may even be spurious. The decline in intensity below $14\ \mu$ could be due to loss of short tracks in scanning and failure of low energy particles to escape from the target, together with multiple scattering in the target. The rise in intensity from $\sim 20\ \mu$ probably represents the beginning of a group of close levels in ^{27}Al above $3.5\ \text{MeV}$. The observed group is much too broad for a single level and is also too intense. For instance, a proton of $\sim 1\ \text{MeV}$ has less than a 2 per cent. chance of penetrating the Coulomb barrier of ^{28}Si compared with a group (2) proton.

There is a further complication in this region of the spectrum. The $^{27}\text{Al}(p, \alpha)^{24}\text{Mg}$ reaction which has a $Q \simeq 1.5\ \text{MeV}$ in proceeding to the ground state of ^{24}Mg will yield α -tracks in the region below $\sim 25\ \mu$. The escape of a $6.2\ \text{MeV}$ α -particle from ^{28}Si is much more probable than that of a 1 or $1.5\ \text{MeV}$ proton. However, grain density measurements (comparing the tracks with thorium α -prongs which occur naturally in every emulsion) favoured protons as the major source of group (6).

The writer is indebted to Professor A. L. Hughes of Washington University, St. Louis, U.S.A., for making this work possible.

References

- DYER, A. J. (1952).—*Aust. J. Sci. Res.* A 5 : 104.
MATHER, K. B. (1951).—*Phys. Rev.* 82 : 133.
REILLEY, E. M., ALLEN, A. J., ARTHUR, J. S., BENDER, R. S., ELY, R. L., and HAUSMAN, H. J. (1952).—*Phys. Rev.* 86 : 857.

INDEX

	PAGE		PAGE
Air Showers, Angular Distribution of High Energy Electrons in	543	Bracewell, R. N., and Roberts, J. A.—	
Air Showers, Residual Range of Delayed Particles in Extensive Alloys, Lattice Component of the Thermal Conductivity of Metals and	410	Aerial Smoothing in Radio Astronomy	615
Aluminium, Scattering of 4.9 MeV Protons by ²⁷ Al ..	57	Bremsstrahlung, Polarization of	527
Anchor Ring, Capacitance of an	658	Brett, J. F.—	
Auroras, Pulsating and Flaming, Association with Complete Ionospheric Absorption at Macquarie Island ..	347	<i>See</i> Garrod, R. I.	77
Auroras, Variation of Intensity of, at Macquarie Island ..	471	Brinkley, T. A.—	
	477	<i>See</i> Titterton, E. W.	350
Baird, K.—		Brisbout, Francoise A., and Hopper, V. D.—	
High Multiple Radio Reflections from the F ₂ Layer of the Ionosphere at Brisbane	165	Delayed Disintegration of Heavy Fragments	352
Ball, F. K.—		Builder, G., and Haneman, D.—	
Sea Surface Temperatures	649	The Noise generated in a Coil with a Ferromagnetic Core	654
Bird, J. R., and Hines, K. C.—			
The Multiple Scattering of Protons in Nuclear Emulsions	586	Calcium, Some Electron Collision Cross Sections of CaII	22
Blatt, J. M.—		Campbell, J. G., and Boyle, A. J. F.—	
<i>See</i> Kalos, M. H.	543	The Response of a Sodium Iodide Scintillation Counter to 18 MeV γ -Radiation ..	284
Bolton, J. G., Stanley, G. J., and Slee, O. B.—		Capacitance of an Anchor Ring	347
Galactic Radiation at Radio Frequencies. VIII. Discrete Sources at 100 Mc/s between Declinations +50° and -50°	110	Carbon, Decay of the 7.68 MeV State in ¹² C	279
Bolton, J. G., Westfold, K. C., Stanley, G. J., and Slee, O. B.—		Chromospheres, Hydrogen, Emission of Radiation from Model	574
Galactic Radiation at Radio Frequencies. VII. Discrete Sources with Large Angular Widths	96	Clothier, W. K.—	
Boyle, A. J. F.—		<i>See</i> Ogilvie, G. J.	210
<i>See</i> Campbell, J. G.	284	Computer, Programme Design for the C.S.I.R.O. Mark I ..	485
		Computers, Effect of Interpretive Techniques on Functional Design of	505
		Conductivity, Electrical, of Monovalent Metals	70
		Conductivity, Thermal, Lattice Component of, of Metals and Alloys	57
		Conductivity, Thermal, of Monovalent Metals	70
		Conductivity, Thermal, of Pure Metals at Low Temperatures according to the Free Electron Theory	64

	PAGE		PAGE
Convection from a Large Horizontal Surface	176	Electrons, Angular Distribution of High Energy, in Air Showers	543
Cornish, E. A.—		Electrons, Multiple Scattering of	217
On the Secular Variation of Rainfall at Adelaide ..	334	Fényes, On a Theory due to ..	14
The Multivariate <i>t</i> -Distribution associated with a Set of Normal Sample Deviates ..	531	Field Theory, Unified, Interpretation of the Field Tensor in	1
Coronal Temperature, The Continuous Radiative Absorption Cross Section of FeXIV and the	25	Filters, Birefringent, Optical Properties of Components for ..	254
Corrigendum	364	Garrod, R. I., Brett, J. F., and Macdonald, J. A.—	
Cosmic Rays, Barometer Coefficient and Air Mass Effects on, at Macquarie Island ..	315	X-ray Line Broadening and Pure Diffraction Contours ..	77
Cosmic Rays, Measurement of Neutron Rate	601	Gas Column, Ionized, Dipole Resonant Modes of ..	268
Counter, Scintillation, Response of a Sodium Iodide, to 18 MeV γ -Radiation	284	Giovanelli, R. G.—	
Crystals, Ionic, On the Criterion for Dielectric Breakdown in ..	36	The Attenuation of Light by Meteoric Dust in the Upper Atmosphere	641
Dielectric Breakdown, On the Criterion for, in Ionic Crystals ..	36	<i>See</i> Jefferies, J. T.	574
Dielectrics, Influence of the Cathode Material on Measured Breakdown Strengths of Solid and Liquid	400	Giovanelli, R. G., and Jefferies, J. T.—	
Diffraction Contours, Pure, X-ray Line Broadening and ..	77	On the Optical Properties of Components for Birefringent Filters	254
Diffraction, Optical, Effects Produced by Amplitude and Phase Changes in the Wave Front	389	The Albedo for the Atomic Scattering of Optical Radiation	570
Disintegration, Delayed, of Heavy Fragments	352	Glasses, Choice of, for Cemented Aplanatic Doublets ..	244
Dungey, J. W., and Loughhead, R. E.—		Godfrey, G. H.—	
Twisted Magnetic Fields in Conducting Fluids ..	5	Optical Diffraction Effects Produced by Amplitude and Phase Changes in the Wave Front	389
Dyer, A. J.—		Haneman, D.—	
The Integral and Differential Range Spectra of Sea-level Mesons	49	<i>See</i> Builder, G.	654
Eccles, P. J.—		Heat Transfer, Vertical, from Impressed Temperature Fluctuations	202
<i>See</i> Officer, V. C.	410	Higgins, C. S.—	
Electron Collision Cross Sections of CaII	22	<i>See</i> Shain, C. A.	130
Electron Screening and Thermo-nuclear Reactions	373	Higgins, C. S., and Shain, C. A.—	
		Observations of Cosmic Noise at 9.15 Mc/s	460
		Hill, G. W.—	
		<i>See</i> Pearcey, T.	485, 505
		Hindman, J. F.—	
		<i>See</i> Kerr, F. J.	297

	PAGE		PAGE
Hines, K. C.—		Kerr, F. J., Hindman, J. F.,	
<i>See</i> Bird, J. R.	586	and Robinson, B. J.—	
Hopper, V. D.—		Observations of the 21 cm	
<i>See</i> Brisbout, Francoise A. . .	352	Line from the Magellanic	
Hopper, V. D., Lim, Y. K.,		Clouds	297
and Walters, Madeline C.—		Klemens, P. G.—	
The Measurement and Red-		The Contribution of Phonons	
uction of Distortion in		to the Thomson Coefficient	520
Thick Emulsions	288	The Electrical and Thermal	
Hydrogen, the 21 cm Line of,		Conductivities of Mono-	
from the Magellanic Clouds . .	297	valent Metals	70
Ionosphere, High Multiple		The Lattice Component of the	
Radio Reflections from the		Thermal Conductivity of	
F ₂ Layer of the, at Brisbane	165	Metals and Alloys	57
Ionospheric Absorption, Com-		The Thermal Conductivity of	
plete, at Macquarie Island,		Pure Metals at Low Tem-	
Association of Pulsating and		peratures according to the	
Flaming Auroras with	471	Free Electron Theory	64
Iron, Continuous Radiative		Light, Attenuation by Meteoritic	
Absorption Cross Section of		Dust in the Upper Atmos-	
FeXIV	25	phere	641
Jacka, F.—		Lim, Y. K.—	
Variations of Intensity of the		<i>See</i> Hopper, V. D.	288
Aurora at Macquarie Island	477	Lithium, On the Photodisin-	
Jacklyn, R. M.—		tegrations ⁶ Li(γ ,d) ⁴ He and	
The Barometer Coefficient and		⁶ Li(γ ,t) ³ He	350
Air Mass Effects on Cosmic		Loughhead, R. E.—	
Rays at Macquarie Island	315	<i>See</i> Dungey, J. W.	5
Jefferies, J. T.—		Ma, S. T.—	
Some Electron Collision Cross		On the Coulomb and	
Sections of CaII	22	Hulthén Potentials	365
Jefferies, J. T.—		Macdonald, J. A.—	
<i>See</i> Giovanelli, R. G.	254, 570	<i>See</i> Garrod, R. I.	77
Jefferies, J. T., and Giovanelli,		Magellanic Clouds, Observations	
R. G.—		of the 21 cm Line from the . .	297
The Emission of Radiation		Magnetic Fields, Twisted, in	
from Model Hydrogen		Conducting Fluids	5
Chromospheres. II	574	Major, G.—	
Kalos, M. H., and Blatt, J. M.—		The Association of Pulsating	
The Angular Distribution of		and Flaming Auroras with	
High Energy Electrons in		Complete Ionospheric Ab-	
Air Showers. I. Landau		sorption at Macquarie	
Approximation	543	Island	471
Keane, A.—		Makinson, R. E. B., and Slade,	
An Investigation of Finite		D. M.—	
Strain in an Isotropic		Dipole Resonant Modes of an	
Material subjected to		Ionized Gas Column	268
Hydrostatic Pressure and		Martyn, D. F.—	
its Seismological Applica-		Comments on a Paper by E. G.	
tions	322	Bowen entitled "The In-	
		fluence of Meteoritic Dust	
		on Rainfall". 2	358

	PAGE		PAGE
Mather, K. B.—		Nicholson, A. F.—	
Measurements of the Cosmic		On a Theory due to I. Fényes	14
Ray Neutron Rate in the		Noise Generated in a Coil with a	
Himalayas and Australian		Ferromagnetic Core	654
Alps	601	Nuclear Emulsion Plates,	
Scattering of 4.9 MeV Protons		Microinterferometric Exam-	
by ^{27}Al	658	ination of	652
See Muirhead, E. G.	527	Nuclear Emulsions, Measure-	
Matthiessen's Rule, Deviations		ment and Reduction of Dis-	
from, for Cold-drawn Wires	210	tortion in Thick	288
Mesons, Integral and Differ-		Nuclear Emulsions, Multiple	
ential Range Spectra of Sea-		Scattering of Protons in	586
level	49		
μ -Mesons, Momentum Distribu-		Objectives, Telescope, Choice of	
tion and Charge Ratio of, at		Glasses for Cemented	
Zenith Angles in the East-		Aplanatic Doublets	244
West Plane	423	O'Dwyer, J. J.—	
Metals, Electrical and Thermal		On the Criterion for Dielectric	
Conductivities of Monovalent	70	Breakdown in Ionic Crystals	36
Metals, Lattice Component of		The Influence of the Cathode	
the Thermal Conductivity of	57	Material on Measured	
Metals, Pure, Thermal Con-		Breakdown Strengths of	
ductivity of, at Low Temper-		Solid and Liquid Dielectrics	400
atures according to the Free		Officer, V. C., and Eccles, P. J.—	
Electron Theory	64	The Residual Range of De-	
Meteoritic Dust, Attenuation of		layed Particles in Extensive	
Light by	641	Air Showers	410
Meteoritic Dust, Influence on		Ogilvie, G. J., and Clothier,	
Rainfall, Comments on a		W. K.—	
Paper by E. G. Bowen on.		Deviations from Matthiessen's	
1, 2	354, 358	Rule for Cold-drawn	
Mohr, C. B. O., and Tassie,		Wires	210
L. J.—			
The Multiple Scattering of		Parry, J. K.—	
Electrons and Positrons	217	See Moroney, J. R.	423
Moroney, J. R., and Parry,		Particles, Penetrating, Residual	
J. K.—		Range of Delayed Particles	
Momentum Distribution and		in Extensive Air Showers	410
Charge Ratio of μ -Mesons		Pearcey, T., and Hill, G. W.—	
at Zenith Angles in the		Programme Design for the	
East-West Plane	423	C.S.I.R.O. Mark I Com-	
Muirhead, E. G., and Mather,		puter. III. Adaptation of	
K. B.—		Routines for Elaborate	
Polarization of Bremsstrah-		Arithmetical Operations	485
lung	527	Pearcey, T., Hill, G. W., and	
Murray, J. D.—		Ryan, R. D.—	
See Wild, J. P.	439	The Effect of Interpretive	
		Techniques on Functional	
Neumann, J.—		Design of Computers	505
Fluctuations of Long-period		Phonons, Contributions of, to	
Accumulations of Daily		the Thomson Coefficient	520
Rainfall Amounts	522	Photodisintegrations $^6\text{Li}(\gamma, d)^4\text{He}$	
Neutron Rate, Cosmic Ray	601	and $^6\text{Li}(\gamma, t)^3\text{He}$	350

	PAGE		PAGE
Positrons, Multiple Scattering of	217	Reflections, Radio, High Multiple, from the F ₂ Layer of the Ionosphere at Brisbane	165
Potentials, On the Coulomb and Hulthén	365	Roberts, J. A.— <i>See</i> Bracewell, R. N. ..	615
Priestley, C. H. B.— Convection from a Large Horizontal Surface ..	176	Roberts, W. O.— <i>See</i> Rush, J. H.	230
Vertical Heat Transfer from Impressed Temperature Fluctuations	202	Robinson, B. J.— <i>See</i> Kerr, F. J.	297
Protons, Multiple Scattering in Nuclear Emulsions	586	Rowe, W. C.— <i>See</i> Wild, J. P.	439
Protons, Scattering of 4·9 MeV, by ²⁷ Al	658	Rush, J. H., and Roberts, W. O.— Recent Studies of Chromospheric Spicules	230
Quantum Mechanics, On a Theory due to I. Fényes ..	14	Ryan, R. D.— <i>See</i> Pearcey, T.	505
Radiation, Cosmic Radio, Comparison of the Intensities of Cosmic Noise Observed at 18·3 Mc/s and at 100 Mc/s..	150	Salpeter, E. E.— Electron Screening and Thermonuclear Reactions	373
Radiation, Cosmic Radio, Observations of Cosmic Noise at 9·15 Mc/s	460	Sea Surface Temperatures ..	649
Radiation, Cosmic Radio, Observations on the General Background and Discrete Sources of 18·3 Mc/s Cosmic Noise	130	Sen, Hari K.— Space Charge Wave Amplification in a Shock Front and the Fine Structure of Solar Radio Noise ..	30
Radiation, Emission from Model Hydrogen Chromospheres	574	Shain, C. A.— A Comparison of the Intensity of Cosmic Noise observed at 18·3 Mc/s and at 100 Mc/s	150
Radiation, Galactic, at Radio Frequencies	96, 110	<i>See</i> Higgins, C. S.	460
Radiation, Optical, Albedo for the Atomic Scattering of ..	570	Shain, C. A., and Higgins, C. S.— Observations of the General Background and Discrete Sources of 18·3 Mc/s Cosmic Noise	130
γ-Radiation, Response of a Sodium Iodide Scintillation Counter to 18 MeV	284	Slade, D. M.— <i>See</i> Makinson, R. E. B. ..	268
Radio Astronomy, Aerial Smoothing in	615	Slee, O. B.— <i>See</i> Bolton, J. G.	96, 110
Rainfall, Fluctuations of Long-Period Accumulations of Daily	522	Smoothing, Aerial, in Radio Astronomy	615
Rainfall, Influence of Meteoritic Dust on, Comments on a Paper by E. G. Bowen on. 1, 2	354, 358	Spicules, Chromospheric, Recent Studies of	230
Rainfall, Secular Variation at Adelaide	334	Stanley, G. J.— <i>See</i> Bolton, J. G.	96, 110
Reactions, Thermonuclear, Electron Screening and ..	373	Steel, W. H.— On the Choice of Glasses for Cemented Aplanatic Doublets	244

	PAGE		PAGE
Strain, Finite Hydrostatic, in an Isotropic Material ..	322	Titterton, E. W., and Brinkley, T. A.—	
Sun, Fine Structure of Solar Radio Noise	30	On the Photodisintegrations ${}^6\text{Li}(\gamma, d){}^4\text{He}$ and ${}^6\text{Li}(\gamma, t){}^3\text{He}$	350
Sun, Harmonics in the Spectra of Solar Radio Disturbances	439	Uebergang, R. G.—	
Swinbank, W. C.—		The Decay of the 7.68 MeV State in ${}^{12}\text{C}$	279
Comments on a Paper by E. G. Bowen entitled "The In- fluence of Meteoritic Dust on Rainfall". 1	354	Walters, Madeline C.—	
Tassie, L. J.—		See Hopper, V. D.	288
See Mohr, C. B. O. ..	217	Wave Amplification, Space Charge, in a Shock Front ..	30
Taylor, N. W.—		Werner, Alma—	
An Interpretation of the Field Tensor in the Unified Field Theory	1	The Continuous Radiative Absorption Cross Section of FeXIV and the Coronal Temperature	25
t - Distribution, Multivariate, Associated with a Set of Normal Sample Deviates ..	531	Westfold, K. C.—	
Temperatures, Sea Surface ..	649	See Bolton, J. G.	96
Thomas, T. S. E.—		Wild, J. P., Murray, J. D., and Rowe, W. C.—	
The Capacitance of an Anchor Ring	347	Harmonics in the Spectra of Solar Radio Disturbances	439
Thomson Coefficient, Contribu- tion of Phonons to the ..	520	Wires, Cold-drawn, Deviations from Matthiessen's Rule for ..	210
Thornton, B. S.—		X-rays, Line Broadening and Pure Diffraction Contours ..	77
Microinterferometric Exam- ination of Nuclear Emulsion Plates	652		

DATE DUE

OCT 17 1982

PERIODICALS MUST BE RETURNED
TO PERIODICALS DESK ONLY

DEMCO 38-297



3 8198 303 614 489

THE UNIVERSITY OF ILLINOIS AT CHICAGO

